



NOVA

IMS

Information
Management
School

MGI

Mestrado em Gestão de Informação

Master Program in Information Management

Smart Cities: Urban Green Infrastructures

Quantifying green infrastructures benefits and value

Filipa Faria de Miranda Teixeira

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

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

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

SMART CITIES: URBAN GREEN INFRASTRUCTURES

QUANTIFYING GREEN INFRASTRUCTURES BENEFITS AND VALUE

by

Filipa Faria de Miranda Teixeira

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

Co-Advisor: Miguel de Castro Simões Ferreira Neto, PhD

Co-Advisor: Pedro Alexandre Reis Sarmiento, PhD

June 2020

ABSTRACT

Populations have moved to cities and urban areas in recent years, resulting in their exponential growth. These migrations raise both traffic and resources consumption, leading to an increase of CO_2 emissions and contributing on a large scale to climate change. To extenuate the environmental issues adjacent to climate change, strengthening the number of urban trees is essential as they mitigate CO_2 emissions when incorporating carbon in their biomass. Understanding and measuring the extent of the role trees play as a mechanism for CO_2 storage introduces an interesting study that provides information on urban green infrastructures planification. Additionally, attending the relation between carbon storage and carbon value, the trees represent a credit benefit for the cities. The incorporation of tree's species, age, and environment type on the previous analysis could also offer valid inputs to enhance the cities' decision making. With the present project, we will provide an interactive dashboard that complete and supports the current analysis.

KEYWORDS

Smart Cities; Urban Green Infrastructures; carbon storage; carbon value, tree species, urban trees, Viseu.

INDEX

| | |
|--|----|
| 1. Introduction | 1 |
| 2. Literature Review | 3 |
| 2.1. Smart Cities..... | 3 |
| 2.1.1. Viseu City | 3 |
| 2.2. Urban Green Infrastructures | 4 |
| 2.2.1. Cities sustainability | 5 |
| 2.2.2. Green Infrastructure Economic Benefits..... | 6 |
| 2.3. Trees impact | 7 |
| 2.3.1. Carbon storage | 8 |
| 2.3.2. Allometric Equations | 8 |
| 2.3.3. Climatic changes | 9 |
| 2.3.4. Carbon Value | 10 |
| 3. Methodology | 11 |
| 3.1. Resource Data..... | 11 |
| 3.1.1. Data Tools..... | 11 |
| 3.2. Cross-Industry Standard Process for Data Mining (CRISP-DM)..... | 11 |
| 3.3. Measure the carbon storage | 12 |
| 4. Data analysis..... | 14 |
| 4.1. Business Understanding | 14 |
| 4.2. Data Understanding | 14 |
| 4.2.1. Describe Data | 14 |
| 4.2.2. Explore data..... | 14 |
| 4.3. Data Preparation | 15 |
| 4.3.1. Rename variables | 15 |
| 4.3.2. Coherence Checking..... | 16 |
| 4.3.3. Handling missing values | 17 |
| 4.3.4. Handling outliers | 18 |
| 4.4. Modelling..... | 18 |
| 4.5. Data Visualization | 21 |
| 5. Results..... | 22 |
| 5.1. Tree Species..... | 22 |
| 5.1.1. Variables Correlation..... | 25 |
| 5.2. Geographic analysis..... | 30 |

| | |
|--|----|
| 5.3. Type | 31 |
| 5.4. Age and State..... | 33 |
| 5.5. Reporting Dashboards | 34 |
| 5.5.1. Overview..... | 34 |
| 5.5.2. Species..... | 34 |
| 5.5.3. Details..... | 35 |
| 5.5.4. Geographic | 36 |
| 5.5.5. Variables Relation..... | 37 |
| 5.5.6. Information..... | 38 |
| 6. Conclusions..... | 39 |
| 6.1. Limitations and recommendations for future works | 39 |
| 7. Bibliography..... | 41 |
| 8. Appendix..... | 1 |
| 8.1. Exploring data..... | 1 |
| 8.2. Data coherence | 2 |
| 8.3. Handling missing values | 3 |
| 8.4. Data Modeler..... | 4 |
| 8.5. Number of trees and carbon storage by species | 5 |
| 8.6. Handling outliers | 6 |
| 8.7. Diameter at breast height e height equations impact | 10 |
| 8.8. Age and State..... | 12 |

LIST OF FIGURES

| | |
|--|----|
| Figure 1: Trees relative frequency | 22 |
| Figure 2: Number of trees and carbon storage, as a percentage, by species..... | 23 |
| Figure 3: CO2 Value per Specie | 24 |
| Figure 4: Number of trees and CO2 relation | 25 |
| Figure 5: Diameter at breast height and CO2 Value relation | 26 |
| Figure 6: Height and CO2 Value relation | 26 |
| Figure 7: Carbon value in Viseu, by census tract | 30 |
| Figure 8: Number of trees by Type and Age. | 32 |
| Figure 9: CO2 value, by age..... | 33 |
| Figure 10: Age and state relation | 33 |
| Figure 11: Overview report, english version | 34 |
| Figure 12: Species report, english version | 35 |
| Figure 13:Details Report, english version | 36 |
| Figure 14: Geographic report, english version..... | 37 |
| Figure 15: Geographic report, english version..... | 37 |
| Figure 16: Information report, english version | 38 |
| Figure 17: Explorating Data Python Code | 1 |
| Figure 18: Part of Python Script results about Exploring Data | 1 |
| Figure 19: Data coherence | 2 |
| Figure 20: Handling missing values - python code I..... | 3 |
| Figure 21: Handling missing values - python code II..... | 3 |
| Figure 22: Data Modeler | 4 |
| Figure 23: Dbhcm distribution, by species family | 6 |
| Figure 24: Hm distribution, by species family..... | 8 |

LIST OF TABLES

| | |
|--|----|
| Table 1: Interval variables Statistics | 14 |
| Table 2: Absolute and relative frequency of the tree species present in the study area | 15 |
| Table 3: Rename variables correspondence. | 16 |
| Table 4: Coherence Checking rules | 16 |
| Table 5: Missing Values information | 17 |
| Table 6: Allometric Equations by Species Tree | 20 |
| Table 7: Diameter at breast height and Height impact, by species family | 29 |
| Table 8: Measures by locality | 31 |
| Table 9: <i>CO2</i> value and relative frequency by location type..... | 31 |
| Table 10: <i>Dbhcm</i> detection of Outliers | 7 |
| Table 11: <i>Hm</i> detection of Outliers | 9 |
| Table 12: Number of trees by age..... | 12 |

1. INTRODUCTION

Technology evolution happens not only at the companies but also in all society- including cities. There has been an increased recognition that more attention needs to be paid to this area. The new cities approach and strategies have based on technology and data, which contributes to develop and adopt the "smart cities" concept. Considering that nowadays cities have data it's possible to have more information and produce knowledge... In other words, with data and technology, we can understand the habitats, workers, and tourists. Posteriorly, the municipalities may plan, adapt, and manage its policies in a more sustainable way, making better and smarter, making better and more smart decisions.

Cities work a crucial role in social and economic features worldwide and have a considerable impact on the environment. According to the United Nations, we predict that in 2030 more than 60% of the world population will live in cities (United Nations, 2018). Consequently, currently, most resources are consumed in cities worldwide, which contributes to economic development and at the same time to their poor environmental status. Cities account for 65% of global energy demand and 75% of global carbon dioxide emissions (Renewable Energy Policy Network for the 21st Century, 2019). However, when we increase the amounts of energy spent with transports and electricity, the lower the urban density, as proved by the fact that CO_2 emissions per capita drop with the increase of urban areas density (Hammer, 2011).

Towns are the centre of economic, people and the environment. The cities metabolism considers the goods input and the waste output with negative externalities, which means social and economic impact. Cities rely on too many external resources, and they are (and probably will always be) consumers of resource (Albino, Dangelico, & Berardi, 2015) . A recent and anthropocentric approach to urban sustainability, cities should respond to people's needs through sustainable solutions for social and economic aspects (Turcu, 2013).

Nowadays, we consider cities that discovery answers, opportunities and solutions for the citizen challenges and problems. Cities worldwide have started to look for solutions which enable transportation linkages, mixed land uses, and high-quality urban services with long-term positive effects on the economy (Albino, Dangelico, & Berardi, 2015). In that way, our mayors and cities are starting to worry about making and creating lively cities to respond to the population's needs. The big goal is offering more life quality and attract people. Therefore, the cities should be smart and give social, economic, environment and substantially policies for its residents, workers, or tourists.

Urban trees are outstanding sieves for urban pollutants and fine particulates. They can offer food, such as fruits. Spending time near trees improves physical and mental health by increasing energy level and speed of recovery while decreasing blood pressure and stress. Trees properly placed around buildings can reduce air conditioning needs by 30% and save energy used for heating by 20–50% (United Nations, 2016).

With the increase of the urban areas, trees could provide a pivotal measure to understand, and, respond to the urban trees impact on carbon storage in cities, as they stock carbon in their tissue though they grow. Trees absorb carbon dioxide and potentially harmful gasses, such as sulphur dioxide, carbon monoxide from the air and release oxygen. Also, they cool the air, land and water with shade and moisture, thus reduce the heat-island effect of our urban communities. The temperature in urban areas is often 9 degrees heater than in areas with substantial tree cover. Trees can help offset the build-up of carbon dioxide in the air and reduce the " greenhouse effect" (College of Agriculture & Life Sciences, 2019)

However, the ecological consequences of those National urban forest planting policies are still not clear. Besides, these kinds of ecological services are often primarily underestimated, and the potential of urban forest (UF) themselves to mitigate CO_2 emission and climate change is far from being understood (Gaston, Ávila-Jiménez, Edmondson, & Jones, 2013). Hence, the precise and timely estimation of UF carbon storage is essential for cities to create strategic systems to improve urban CO_2 emissions mitigation (Ren & Zhai, 2019).

Increasing flora cover in cities could be a pivotal approach to justify urban temperature excess. One way to measure the interactions between the urban environment and vegetation is to complete microclimate models that can simulate the effects of vegetation onto the urban microclimate, as well as the effects of urban environments onto vegetation. To provide reliable estimates, microclimate models need to be parameterised based on empirically obtained data.

At the same time, Environmental preoccupations are being more significant and are becoming an essential business for all of us. These issues range from global warming, temperature, CO_2 emissions, acidification, melting and can even lead to extreme weather phenomena. Today these aspects are even more relevant, and CO_2 emissions play a central role in the political agenda, due to the recognition that CO_2 emissions are the primary provider to climate change (United Nations , 2019). Attending to the global and government preoccupation, the administration needs information to support its decisions.

The goal of this thesis is to analyse the urban Trees' power to measure carbon storage and Study trees' impact on air quality in urban areas. We will combine the carbon storage with carbon produced and carbon monetary value of Viseu city which could represent a better approach to support the government decision making. For instance, at the beginning, we will address other points of view on the sustainable use of trees. Nevertheless, we found some challenges at the city data level.

To support the decision making and our analysis, we will develop a dashboard using Power BI to better relate the additional variables like location, environment, age and status tree with its number, carbon storage and carbon monetary value. Our analysis will focus on three perspectives: species tree, geographic location and type age and status.

2. LITERATURE REVIEW

2.1. SMART CITIES

"Smart City" concept does not have a standard definition. International Data Corporation (IDC) defines a "Smart City" as a city that has declared its intention to use information and communication technologies to transform *modus operandi* into one or more of the following areas: energy, environment, government, mobility, buildings, and services. Smart cities want to increase their life quality, citizens and implement a sustainable economic growth. At the same time, the modernisation and information knowledge became crucial elements to discover the decisions making. It involves the use of digital and communication technologies and, as such, it strives towards high-quality resource management and service delivery (China Academy of Information and Communications Technology, EU-China Policy Dialogues Support Facility II, 2016). Considering other common concepts, like "digital city" or "city of the future", the difference is in "smart", which means data management and contains as one of its essential features project implementation from the "bottom-up" through the involvement of local communities (Crncevic, Tubic, & Bakic, 2017).

Currently, the leading Smart City focus on habitant/men. The technology evolution became part of the business, companies, and markets. In addition, data was spread all over society as well as the power to transform or convert them into knowledge. We can know what people bought, view, visited, liked or not. With this information, we could look at our cities and analyse social behaviour, to support decision making.

Urban territories contribute roughly seventy-five per cent on non-renewable energy source inferred CO_2 emanations, and numerous urban areas have instituted discharges relief plans. Assessment of the viability of relief endeavours will require estimation of both the discharge rate and its change over the existence (Turnbull, et al., 2018). Cities strengthen population, business, industries, and energy output. Consequently, these are critical focal points to study improvement and emission CO_2 drivers' opportunities.

2.1.1. Viseu City

The city is in Portugal Centre (40° 39'39.64 "N, 7° 54'34.96 "W). The municipality altitude is irregular, and it varies between 400 and 700 metres. The covered area is around 507.10 km^2 and divided into 25 parishes. Viseu is around 50 km east of Atlantic Ocean, encircled by mountains and it crossed by a rivers and streamlets network.

These characteristics make Viseu a city with a microclimate, with extreme temperatures: cold and wet winters and hot and dry summers. The average temperature is 19.2°, and the average precipitation is 1,198.5 mm. On the west side of the city, we can find Serra do Caramulo, which plays a significant role in climatic terms by reducing the impacts of the western air masses (while the Mondego River's basin makes the precipitation easier).

Viseu is also capital of the district with a population of 100 000 inhabitants. Two years ago, Viseu was in the top three of the healthiest councils, with greater security and quality of life (Cardoso, 2018) .

The municipality recognises resident's life quality as one of its priority goals, and it intends to become the best city to stay like to become the best city to stay, get a job, report, and invest (Cardoso, 2018). Viseu seeks to invest in tools that fix people that already live there, as well as attract new residents.

In the environmental field, *Cidade Jardim* project emerged consisting on an exhaustive gathering of the city's trees species. This project focus on a differentiated logic, for instance in the matter of pruning and the respective time of the year, including the analysis of late pruning. That will reduce the effect of pollens on people with allergies, or plan for replanting, removing some species to integrate others, explains Almeida Henriques (municipality mayor).

Additionally, the city has invested in improving the city's mobility infrastructures, with the creation of the new "Viseu Mobility Operations Center" and "Interface", the insertion of postcards, the implementation of the "Viseu Seguro" program" and implementing an environmental education project "Viseu Educating for Sustainable Mobility".

This town, by development all these initiatives, works to respond the goal of reducing more than 96,000 tons/year of CO_2 by 2030, the promotion of energy and environmental sustainability, the promotion of the Information Society and Knowledge and promotion of active and healthy ageing in the county.

Consequently, we believe that the present study may be an essential tool to support the municipality's decision. Thus, we will make it possible to add more information to ongoing projects and contribute to the control of CO_2 emissions per year, specifically (Ferreira, 2020).

2.2. URBAN GREEN INFRASTRUCTURES

After the industrial revolution, the relationship between man and industry increased and became more prominent, causing a more significant influx of people in cities and a decline in agricultural and rural environments. Cities growth process was becoming faster and faster, marking the 20th century with the predominance of the city instead of the countryside.

During the 1920s and 1930s, modernism had invaded the cities. It had been segregated it in large functional areas. There, the focus had been on the economy with large commercial and industrial buildings and, at the same time, housing in big and high buildings as well. The green spaces appeared as multiple spaces for leisure. Since well-being in cities was associated with buildings, there was a contribution to less green spaces and environments.

Shortly after that, transports gave the impression that the city could disperse, as it allows for the intertwining of small towns. However, in practice, with the oil crisis in the 1970s, this concept became unsustainable. Urban expansion had an increasing impact, reaching levels of suburbanization (Mark Scott, 2016).

Then the new Urbanism was born, where compact and close city prevailed, as a result of a vibrant transport network. If, on one hand, rural spaces have been preserved, due to their abandonment, urban spaces have increasingly abdicated green spaces. At the same time, in the most notable towns, the transport network contributed to the reduction of car and the consolidation and greater efficiency in urban infrastructure (Mark Scott, 2016).

However, technological advances have introduced Geographic Information Systems in city planning as a tool for economic analysis and support for urban development. Furthermore, considering that society was becoming more aware of the impacts of the city's environmental on human health and the benefits of green spaces for welfare municipalities were forced to focus on other values of natural areas beyond nature and biodiversity, namely those that would also respond to the needs of the populations.

The need to reverse urban development behaviours and the emergence of the sustainable development concept, in 1987, as a development to seek the satisfaction of the current generation's needs, without compromising the ability for future generations to satisfy their own needs (United Nations, 1987), emphasizes the integration of the environmental component in planning. Thus, Green Infrastructure emerges as a territorial planning instrument, intending to defend the need for a more strategic and comprehensive approach to soil conservation.

Green Infrastructure concept has underlying notions that goes back to the beginning of Environmentalism, Nature Conservation, and the intervention of Landscape Architecture in 19th-century urbanism (Sussams, Sheate, & Eales, 2014). The Green Infrastructures is based on connections between green spaces and parks, in sequence to help populations and biodiversity preservation around the allocation of counterpoint corridors the habitat fragmentation (Samora-Avela, Ferrão, Ferreira, Panagopoulos, & Vaz, 2017).

Accordingly, at a policy level, planning and the planning system needs to incorporate green infrastructures and an ecosystem approach to ensure that benefits are optimized in the long term, especially concerning global warming, biodiversity loss and climate change adaptation (CIWEM, 2010). We must plan and decide the infrastructures, in order to recognise the fundamental responsibility, it plays in the ecological and climate mindful growth of urban areas. Green infrastructures have an extensive range role with impact in quality life, and all the subsequent and related development from the beginning.

By taking care of the urban zones, the regards for the planet expands and green territories become more attractive, which increases the search for multifaceted, valued by both humans and the government. In that way, people have an interest in integrating dynamic and further action from cities, and the green infrastructures into the plan from development cities, becoming the green infrastructures plan a critical goal.

2.2.1. Cities sustainability

Urban Green Infrastructures can concentrate a big range of urban challenges, such as conserving biodiversity, adapting to climate change, supporting the green economy, and improving social cohesion. To capture this potential, local governments need to plan carefully and holistically.

According to Forest Research, there is a bright collection of proof which exhibits the remedial estimation of green space demonstrating that increasingly inactive types of use, or even merely access to perspectives on green space, can beneficially affect mental prosperity and individual capacity (Forest Research, 2010). We could provide opportunities for relaxation, exercise, and interaction with green areas, and when individuals are exposed to a green environment, stress levels decreased fast, compared with people who did not have this opportunity.

Maintainable reactions to economic and climate problems can help to solve social problems, such as being poverty and traffic congestion. Regardless, green spaces can also facilitate social interaction, integration, and the development of community cohesion (Forest Research , 2010). According to the Institute of Public Health reduced access to the natural environment can result in social isolation, obesity, and chronic stress. Consequently, access to green space is a significant predictor of increased physical activity (“active living”) and reduced risk of obesity (Heinze, 2011). Therefore, if Green infrastructures are related to parks, they play an essential role in helping physical activity in minority societies. It has been confirmed that people walking around parks shows a reduced stress rate across a broad spectrum of individuals (Hartig, Davis, Jamner, & Jamner, 2003).

Urban infrastructures and vegetation are typical for ecosystem assistance. They can decrease problems with flooding and generate positive effects on air quality (as a filter of polluted air). At the same time, improvements in air quality due to vegetation have a positive impact on physical health with apparent benefits as a decrease in respiratory illnesses (CemilBilgili & Gökyer, 2012). When people are in contact with green spaces, their stress levels decrease, which makes the correlation between nature and people more significant and with a greater influence on people’s lifestyle, work productivity and mental health.

The health benefits of regular walking have been widely reported, confirming that the benefits of this activity can contribute to reduce coronary heart disease and type 2 diabetes (Jones, M., & Coombes, 2009). Hence, green infrastructure can also support active travel, with incorporated walking and cycling, stimulating health.

Nevertheless, the links between green space and physical activity are not clear, and different studies have found contradictory results (Jones, M., & Coombes, 2009). Sometimes public spaces can attract criminal activities, such as drug dealers, nope – rapists. These criminal activities occur mainly because green spaces are often deserted at night, providing secluded and convenient venues for crime (Z., 2013). Nevertheless, a public study development in an inner-city found a correlation between lower crime rates and nearby vegetation (Kuo, 2001). That way, natural areas had the power to reduce crime and increase productivity, promote psychological wellbeing, decrease stress, and improve immunity.

The effects in urban environments, cannot be over-stated and sometimes green space benefits are disproportionately distributed through the society. In other words, those who live in weak areas, the minority, the elderly, women, and people with disabilities, may not have access to the same levels of benefits. However, when the network has engaged with the arranging and usage of a green territory, its essence could be expanded as it assumes control over administration and responsibility for green territories in their region (Z., 2013).

2.2.2. Green Infrastructure Economic Benefits

Still, Green Infrastructure offers a high-quality environmental setting attracting new businesses which directly serve tourism, recreation, leisure, and health sectors (Forest Research , 2010). Frequently, people develop local activities where they live which contributes to urban green spaces becoming an input for relax and a local to usufruct these activities. However, the benefits are not only concentrated on the inhabitants of each region. Green infrastructure can attract and develop tourism, new local shops and get ideal exteriors for sports activities and a recreational variety.

Green infrastructure has also used as a valuable education resource. It has the potential to improve educational achievement, eventually helping to create a better qualified and more highly skilled workforce and bringing higher salaries and more valuable business investment into the region (Greater London Authority (GLA), 2008). With these green infrastructures in the cities there's an increase in opportunities for educational events and jobs as forest warden, which consequently, creates more occasions to develop learning and employment.

The green infrastructure promotion and better respect can help decreasing management costs, like reduction in litter and other inappropriate practices. Green infrastructure provides a valuable, local resource for local schools and parents to use as a way of teaching children about sustainable development, conservation, environmental change, places, plants, animals, and natural and human-made processes (Okunlola, 2013). So, green infrastructure is a mechanism for sensitize and tell people about their environment.

Investment in Green Infrastructure can inspire and attract high-value industry, entrepreneurs, and skilled workers to a region through the creation of high quality, environmentally friendly living and working environments, value-adding to local economies (ECOTEC, 2008). It can offer leisure and recreation, opportunities creation and stimulate economic activity (forestry, parks, agriculture, and public services) with green infrastructure investment. These places are a way to attract and develop high-value industries, start-ups, businesspersons, and employees. Thereby, increasing the green infrastructures quality, the impact on the environment and society can also be positive so that it may add to economic performance.

When cities increase their green economy, there are also linked prospects. They can create new jobs, limit the environmental impact of towns and cities, and reduce the costs of running them (Sustainable development Commission, 2010). Also, considering the minority, food acquiring makes up a significant part of a household's income. So, the production that results from urban agriculture can be used for home consumption and as an effective way of supplementing income, thus contributing to poverty reduction (Z., 2013) . In this way, attending to subsistence, green infrastructures contribute to less expensive and capital cost, increasing the cost-effective approaches to manage runoff municipalities.

The urban public park can help in increasing the property value the real estate market consistently demonstrate that many people are willing to pay a higher amount for a property located close to parks or open space areas than for homes that do not offer this facility (Okunlola, 2013). A park becomes one of a city's innovations and attraction creation as a crucial advertising instrument to invite business, tourists, and conventions. Although the economic signal for the delivery of these profits is imperfect, these parks and green spaces can become profitable. The new green infrastructure development of current and new investments should be prioritized, particularly in areas that are significantly in need.

2.3. TREES IMPACT

Due to the growth of cities and harmful human activities, some species lost them habitats. At the same time, considering the environment and the species, this fact can become an opportunity to preserve some species. Nowadays, many cities contain sites of particular importance for conservation because they protect threatened species and habitats (CBO, 2012). So, many native fragments of vegetation

survived because of their geography, earth, and other characteristics. Increasing and planning cities according to the urban green infrastructures, could contribute to the biodiversity perseverance and control of some species behaviour.

Urban trees play a fundamental role in CO_2 storage as they act as a sink for CO_2 by fixing carbon during photosynthesis and storing excess carbon as biomass (Nowak & Crane, 2002). That way, we need to understand the power and impact of the urban trees on carbon storage, as they collect carbon in their mass while growing.

Trees cool the air by throwing conceal and discharging water fume trees can channel fragile issues of particles - one of the most dangerous types of air contamination, produced from copying biomass and oil derivatives (The Nature Conservancy, 2016). On the other hand, to face the amount of carbon released into the atmosphere, derived from cities' dimensions and human activities, the growth of trees has a greater importance and impact. Considering that the emissions caused by humans are higher than the number of trees saved, urban trees cannot be the only moderates. In parallel, trees also contribute with carbon to the atmosphere- when they die and rot. Consequently, by understanding how many carbons are stored by trees, we can become conscious about the amount of carbon that could be released when trees cease to exist.

Tree planting is becoming a strategy for city planners to mitigate heat and air pollution. Nevertheless, only trees can both renew and clean the air. Besides, trees and other green framework give a broad scope of co-benefits too—including territory for untamed life, stormwater control, entertainment openings, and beautification of open and private spaces in urban communities (The Nature Conservancy, 2016).

2.3.1. Carbon storage

In their biophysical process, trees trap and release CO_2 into the atmosphere. Throughout photosynthesis, through leaves and by using the sun's energy, trees produce oxygen, vitamins, resins, and hormones necessary for their growth and health. It is based on the Carbohydrates retained throughout photosynthesis that trees grow, from where the respiration originates CO_2 storage by the trees.

According to Aguaron and McPherson, the term "carbon dioxide storage" refers to the accumulation of woody biomass as trees grow over time. The similarity amidst the tree's biomass and the CO_2 stored amount by it, in urban trees case, is proportional and it impacts on the tree density and management procedures (McPherson, Peper, & Van Doorn, 2016).

2.3.2. Allometric Equations

The allometric equations are the base to compute the carbon storage, using the diameter at breast height (Dbh), tree height, wood density, moisture content, site index and tree condition. Some of these parameters are correlated with climatic status and geographic area. Hence, some error is related to the use of regular densities and humidity contents in allometric formulas. We have two allometric biomass equations forms: volumetric and direct. Volumetric equations compute the overhead ground

volume of a tree using *Dbh* and tree height for the species. Linear equations produce beyond the ground dry weight of a tree using *Dbh* and tree height (McPherson E. A., 2012).

When we use allometric equation, we increase the uncertainty of our study. Though we have a lot of allometric equations developed for city trees, their precision needs to be more established when we apply it on an area with a challenging climate to define and to trees with significant size oscillations (McPherson E. A., 2012).

In city areas, we usually find many different tree species. Attending to E. Aguaron and, E.G. McPherson, we can find around 100 species in a city. Considering these tree diversity species and the allometric equation types scarcity for urban areas, most of the trees will be associated to a similar species or a family species equation. This fact increases our computation error associated with the equation choice (McPherson E. A., 2012).

2.3.3. Climatic changes

The Earth's climate has transformed during history. Few Scientists show an Earth-wide temperature boost propensity. This was seen since the mid-twentieth century to the human extension of the "nursery impact" — warming that outcomes when the climate traps heat transmitting from Earth toward space (NASA, 2018).

With climatic changes arises the Global warming concept, as a long-term rise in the average temperature of the Earth's climate system. It has been confirmed by direct temperature measurements and by various effects of warming, and its importance for climatic changes getting higher and higher (NASA, 2010). Some gases presented in the atmosphere are blocking the heat escaping. Long-lived gases stay "permanently" in the atmosphere "forcing" climate change (Stocker, et al.). With human activity, especially pollution and fossil fuels exploration, the temperature increases. Higher world temperature provides specific environmental problems, like melting of the ice caps, rising sea levels, acidification of the oceans, destruction of ecosystems, extinction of species, and, at least causes increasingly intense and frequent extreme weather phenomena (NASA, 2010).

On Earth, human activities are altering natural climatic. Over the last century, fossil fuels consumption of coal and oil has increased the intensity of atmospheric carbon dioxide (CO_2). With the consumption of coal or oil, oxygen associates with carbon that is discharged into the air, shaping the CO_2 . To a lesser degree, the freeing from land for horticulture, industry, and other human exercises has expanded convergences of ozone-depleting substances (NASA, 2010). Considering the impact of the trees, they could play a critical role in the climatic changes combat.

The climate changes cause impacts on our entire society (business, families, and governments). To control it, we have built the social cost of carbon as a measure of the economic harm from those impacts.

2.3.4. Carbon Value

As we already mentioned, the carbon value corresponds to the social cost of carbon. It is expressed as the amount for the total damages from emitting one ton of carbon dioxide into the atmosphere, in euros (Environmental Defense fund, s.d.).

Attending to the Health Economic Assessment tool approach the carbon monetary value changes by each country and year. The price has in consideration the international facts, regional and country values.

When we compute the carbon monetary value of carbon, we calculate the hypothetical credits generated by the urban areas. This value, at the same time, stimulates climate-friendly securities, contributing to climate safety (Stempski, 2018).

3. METHODOLOGY

In this division, we introduce the methodology applied. First, we refer to the data source and data type used. Second, we present the data tools, techniques, and software used. Subsequently, we explain the data process adapted to respond to our problem since the input validation process data come out of the transformations and models required to final data set. Additionally, we describe the model methodology adopted, i.e., the formulas behind the request calculations.

3.1. RESOURCE DATA

The study is centred on the analysis and visualization of the secondary data obtained from the partnership between *NOVA Cidade* and *Câmara Municipal de Viseu* (Viseu City council). The data set was collected by the council government to develop studies and smart city projects to support decision-makers.

The project goal attends to quantify the carbon storage value and quantity using the most common tree species of Viseu.

3.1.1. Data Tools

This project mainly focuses on the Power BI tool. In the beginning, the dataset made available was in XML format. It was imported from Microsoft Excel to Microsoft Power BI.

In the initial part data processing and pre-processing and calculation of barometric equations, we will be using Python scripts inside Power BI. Later, when we create dashboards and make a visual data presentation, we will use Data Analysis Expressions (DAX).

3.2. CROSS-INDUSTRY STANDARD PROCESS FOR DATA MINING (CRISP-DM)

Cross-Industry Standard Process for Data Mining (CRISP-DM) is a six-stage model of the whole DM process. From beginning to end, that is exhaustively material across undertakings for a far-reaching display of data mining adventures. This methodology transforms business problems into data mining tasks, permitting holding projects independent of technology and applications (Wirth, 2000).

In the beginning, the CRISP-DM starts with a **business understanding** step. Here, the project goals are specified. This step determines a problem. At the “**Data Understanding**” phase, we need to describe the data, explore data, and verify data quality for the next steps in order to have the right data for our problem already defined. Afterwards, it follows the “**Data Preparation**” phases, when we analyse data statistic, missing values, inconsistent data... the data focus on the business goal and adapts the values to that propose.

Additionally, the data sets can be “**Modelling**” to discover some parameter settings as desired, we built the data mining workflow for the nominated algorithms and to complete the data mining task on the

pre-processed data. In this phase, we adopt allometric equations to measure carbon storage. Surrounded by “**Evaluation**” the results reviewed attending to the core business objectives. Last, but not least, follows the “**Deployment**” step, where we implement or publish the project developed.

We will use the CRISP-DM methodology to develop this project. First, we will clarify the project goals, understand, and prepare the data, with python scripts inside Power Bi. The modelling phases will require Allometric Equations, explained above. In the end, we respond to the evaluation and the deployment phases with Power Bi, building a dashboard to respond to the problem and to understand the results better.

3.3. MEASURE THE CARBON STORAGE

In order to measure the carbon storage (C_n) of each tree species, first, we started with the calculation of the trees' volume. To discover the volume value, we adapted the allometric equation for each tree, attending to its details and using the height and diameter of the tree Top (DAP_cm and $Altura_m$) variables. After that, we calculated the dry weight biomass, multiplying the weight density factor (specific value for each specie) (Neto & Sarmiento, 2019) by the volume already calculated. Later, to calculate the total weight biomass and carbon and stored carbon dioxide, we multiplied the dry weight biomass (Dwb_n) by constant values. First, we computed the total dry weight biomass, the previous dry weight biomass value plus the belowground biomass, with the product between Dwb and 1,28 (Husch, Beers, & Kershaw, 2003). At last, we converted the Twb to Dwc , multiplying by $\frac{1}{2}$ (Leith, 1963). Ultimately, we estimated the stored carbon with the product between $\frac{44}{12}$ and Dwc (McHale, 2009).

So, in short, with the purpose of measure the Viseu trees C_n , we calculated:

1. The trees volume by applying allometric equations according to the previous conditions:

$$Vt_n = f(\text{diameter at breast height, height});$$

2. The trees dry weight biomass, i.e., the biomass of the tree visible part. For each tree was considered a certain dry weight density, according to the tree species, adopting the formula:

$$Dwb_n = Vt_n \times Df_n, \text{ with } Df_n = \text{weight density factor};$$

3. Total dry weight biomass: part of the over and underground:

$$Twb_n = 1.28 \times Dwb_n;$$

4. Dry weight carbon, that is, allows us to determine the carbon absorbed value for each tree:

$$Dwc_n = Twb_n \times \frac{1}{2};$$

5. Stored carbon: multiply the Dwc by $\frac{44}{12}$:

$$C_n = \frac{44}{12} \times Dwc_n.$$

After we calculated the stored carbon (C_n), we considered relevant calculated carbon monetary value associated. For each tree of Viseu, based on molecular weight, we compute the carbon monetary value with the product between the molecular weight of CO_2 -3.67- and carbon quantity. The CO_2 value for each tree multiplied by the carbon projected social cost for Portugal in 2019: $\frac{12.74\text{€}}{tCO_2}$ (World Bank Group, 2019). The carbon cost social value of this tool is presented by country and year, based on global evidence, local averages, or country-specific values.

4. DATA ANALYSIS

4.1. BUSINESS UNDERSTANDING

The project aim is to study the potential of Viseu urban trees to measure stored carbon and carbon monetary value. For such, we had to estimate Vt , Dwb , Twb , Dwc , and C . At last, we provided a dashboard presentation to relate the variables and provide perspectives for analysis results.

4.2. DATA UNDERSTANDING

4.2.1. Describe Data

The dataset provides a total of 12,579 observations and 66 variables and corresponds to data collected between February 2017 and June 2018. In the beginning, we selected the most relevant variables for the analysis. We eliminated the ID variables and variables with higher missing data. To select which variables, we had to include in the analysis we built a Python script that provides a descriptive analysis, identifying the quantity observation and data type for a variable, [see Appendix 1].

4.2.2. Explore data

4.2.2.1. Interval Variables

After describing the data, we provided a statistic analysis of the interval variables- minimum, maximum, average, count and standard deviation for the variables, *table 1*.

| Variable | Min | Max | Average | Count | Standard Deviation |
|-------------------------------------|-----|-------|---------|--------|--------------------|
| Tree branch height (m) | 0 | 840 | 2.88 | 12,579 | 14.73 |
| Diameter at breast height (cm) | 0 | 227.5 | 22.03 | 12,579 | 19.85 |
| Circumference at breast height (m) | 0 | 29.30 | 5.39 | 12,579 | 3.72 |
| Circumference at breast height (cm) | 0 | 2930 | 538.82 | 12,579 | 372.30 |
| Height (m), | 0 | 63 | 9.03 | 12,579 | 6.35 |

Table 1: Interval variables Statistics

4.2.2.2. Categorical Variables

For the categorical variables, we calculated the number of individuals and their relative proportions for each category. With the presented analysis, we observed that the dataset provides data about a lot of scientific species. So, we focused our study on 20 most representative species, *table 2*:

| Scientific Name | Nr. Of Trees | Relative Frequency |
|--------------------------------|--------------|--------------------|
| <i>Acer negundo</i> | 1,337 | 14.32% |
| <i>Liquidambar styraciflua</i> | 912 | 9.77% |
| <i>Tilia tomentosa</i> | 682 | 7.31% |
| <i>Quercus robur</i> | 637 | 6.82% |
| <i>Fraxinus excelsior</i> | 594 | 6.36% |
| <i>Ligustrum japonicum</i> | 486 | 5.21% |
| <i>Populus nigra 'italica'</i> | 480 | 5.14% |
| <i>Platanus x hispanica</i> | 450 | 4.82% |
| <i>Caldeira vazia</i> | 435 | 4.66% |
| <i>Tilia cordata</i> | 435 | 4.66% |
| <i>Tilia platyphyllos</i> | 396 | 4.24% |
| <i>Celtis australis</i> | 378 | 4.05% |
| <i>Acer pseudoplatanus</i> | 329 | 3.52% |
| <i>Fraxinus angustifolia</i> | 313 | 3.35% |
| <i>Cercis siliquastrum</i> | 297 | 3.18% |
| <i>Aesculus hippocastanum</i> | 275 | 2.95% |
| <i>Acer saccharinum</i> | 247 | 2.65% |
| <i>Betula pendula</i> | 226 | 2.42% |
| <i>Betula celtiberica</i> | 213 | 2.28% |
| <i>Melia azedarach</i> | 213 | 2.28% |

Table 2: Absolute and relative frequency of the tree species present in the study area

4.3. DATA PREPARATION

On the Data Preparation, we started by renaming some variables, attending to the names provided by the methodology adopted. After that, we checked the incoherencies in numerical variables, handling missing values.

4.3.1. Rename variables

The *table 3* gives the new name variables that we had adopted.

| Variable | Variable source name | New name |
|-------------------------------------|----------------------|------------|
| Tree branch height (m) | Alt_ramo_m | hb_m |
| Diameter at breast height (cm) | DAP_cm | dbh_{cm} |
| Circumference at breast height (m) | d_copa_m | dc_m |
| Circumference at breast height (cm) | d_copa_cm | dc_{cm} |
| Height (m), | Altura_m | h_m |

Table 3: Rename variables correspondence.

4.3.2. Coherence Checking

The coherence checking helped to verify if the analysis dataset had consistent data. Some rules were checked, to see if there was any incoherence data. Furthermore, in the beginning, we checked if the dataset had duplicates, although not applicable. For that, conditions were created to delete the incoherence data, *table 4*:

| Variable | Rule | Explanation |
|--------------------|-----------------------------|--|
| dbh_{cm} | $hb_m > 0$ | All variables should assume values higher than zero. |
| dc_m | $dbh_{cm} > 0$ | |
| dc_{cm} | $dc_m > 0$ | |
| h_m | $dc_{cm} > 0$ | |
| hb_m | $h_m > 0$ | |
| dc_m & dc_{cm} | $dc_{cm} \times 100 = dc_m$ | For convert meters to centimetres we multiply the distance value, in <i>cm</i> , by 100. |

Table 4: Coherence Checking rules

Additionally, we removed the duplicates and error values, in total, we deleted 435 observations, corresponding to a proximally 3% of the original data set. The summary of this python script can be seen in the [see Appendix 2].

4.3.3. Handling missing values

The variables with missing values are categorical variables. So, to handle missing values, we applied an Extra-Trees Regressor and we filled the missing values with predictive values. This technique implements a meta estimator that fits randomized decision trees on various sub-samples of the dataset. It uses averages to adjust the predictive precision and control over-fitting.

The decision trees split the data into sets of mutually restricted records. When we used more than one decision tree, we built a decision forest that identifies data with even better similarity (among the records) and higher correlations (among the attributes). We used this technique because sometimes, a dataset probably has records within a section that have higher similarity and attribute correlations comparing to the similarity and correlations of the whole data set (Rahman & Zahidul, 2013). Additionally, when we use the data partition for the replacement, we can give better accuracy imputation.

With Extra-Trees Regressor, a random dataset partition is applied, however, it extracted the best random thresholds. At the same time, these are randomly selected the same as the splitting rule. It permits reducing the model variance, increasing the bias.

Once we had categorical variables, we had to transform it in numerical variable. In that way, we created an index that corresponds each categorical variable to a number. Attending to the variables used, we found a total of 6 variables with missing data, *table 5*:

| Variables | Missing Values |
|----------------------|-----------------------|
| Idade | 1387 |
| Envolvenc | 1304 |
| Projecao_c | 1310 |
| Anal_Local | 2033 |
| Estado | 777 |
| Missing Values Total | 6811 |

Table 5: Missing Values information

After that we built a python code, [see Appendix 3], to apply the predictive decision tree imputation. Once we are using categorical variables this imputation it could be influential in correlation between the variables. With the python code our dataset had numerical variables rather than categorical and text variables. So, to impute the values in really dataset, we merged by OBJECTID variable the data set before and after imputation.

At this point we had a new dataset like an outer join of the previous table and the new table-without the missing values. First, we built support tables with index and respective descriptions and after, we connected the additional tables with the bases. So, inside Power BI we built the correspondence between the new numerical values, that we produced applying the decision tree technique, and its description, presented in additional tables.

Additionally, we found missing values on variables *freguesia* and *localidade* (parish and zone). Heretofore, we have the street, latitude, and longitude of each tree, manually we substituted the missing values for the correct value.

4.3.4. Handling outliers

The current study focusses principally on descriptive statistics analysis. Attending to that, we will handle the outliers just for the study part of dimension tree variables and it is the correlation with CO_2 .

4.4. MODELLING

To compute the new variables (volume, total dry weight biomass, total dry weight biomass of carbon, dry weight biomass, and carbon storage), we used allometric equations that we recovered from the GlobAllomTree database (Henry, et al., 2013) presented on the website (<http://globalometree.org/>). Since not all trees have an allometric equation associated, but only one similar tree species, which in turn has an allometric equation, we grouped the species tree in family species. When there were no similar species of the equivalent genus, we applied a general equation for broadleaf trees (Mcpherson & Doorn, 2016). In our equations, we considered the Dbh_{cm} and the h in meters, *table 6*.

| Tree species | Allometric equation | Family species | Source |
|------------------------------|--|----------------|--|
| <i>Acer Negundo</i> | | | |
| <i>Acer Pseudoplatanus</i> | $V = 0.012668 + 0.0000737 \times dbh^2 \times h^{0.75}$ | Aceraceae | (Centre for Ecology & Hydrology, 2005) |
| <i>Acer Sacharinum</i> | | | |
| <i>Tilia Cordata</i> | | | |
| <i>Tilia Platyphyllos</i> | $V = 0.00004124 \times 10^{(1.9302 \times \log dbh) + (0.0209 \times (\log(dbh))^2) + (0.129 \times \log h) - (0.1903 \times (\log h)^2)}$ | Malvaceae | (Giurgiu, 1974) |
| <i>Tilia Tomentosa</i> | | | |
| <i>Fraxinus Excelsior</i> | | | |
| <i>Fraxinus Angustifolia</i> | $V = 0.00030648 \times 10^{1.2676 \times \log_{10}(dbh) + 0.3102 \times \log_{10}(dbh)^2 + 0.4929 \times \log_{10}(h) + 0.0962 \times \log_{10}(h)^2}$ | Oleaceae | (Giurgiu, 1974) |
| <i>Betula Pendula</i> | | | |
| <i>Betula Celtiberica</i> | $V = 8.141 \times 10^{-5} \times 10^{2.248 \times \log(dbh) - 0.2062 \times \log(dbh)^2 + 0.1946 \log(h) + 0.4147 \log(h)^2}$ | Betulaceae | (Giurgiu, 1974) |

| Tree species | Allometric equation | Family species | Source |
|--------------------------------|---|------------------|---------------------------------------|
| <i>Quercus Robur</i> | $V = 0.00009464 \times 10^{1.9341 \times \log dbh - 0.0722 \times \log (dbh)^2 + 0.6365 \times \log h + 0.172 \times \log (h)^2}$ | Fagaceae | (Giurgiu, 1974) |
| <i>Platanus X Hispanica</i> | $V = 0.0000485 (dbh^{2.43642}) (h^{0.39168})$ | Platanaceae | (Pillsbury, Reimer, & Thompson, 1998) |
| <i>Celtis Australis</i> | | | |
| <i>Cercis Siliquastrum</i> | | | |
| <i>Melia Azedarach</i> | | | |
| <i>Liquidamba Styraciflua</i> | $V = 0.0001967 (dbh^{1.951853}) (ht^{0.664255})$ | General Equation | (McPherson, Peper, & Van Doorn, 2016) |
| <i>Ligustrum Japonicum</i> | | | |
| <i>Populus Nigra 'Italica'</i> | | | |
| <i>Catalpa Bignonioides</i> | | | |
| <i>Other Species</i> | | | |

Table 6: Allometric Equations by Species Tree

Attending to higher quantity of data and species trees, we considered at least 75% and we found an allometric equation or apply the general equation if it is the case. The other 25% of the trees assume the general equation too.

After that, we computed the additional variables (*dwb*, *Twb*, *Dwc*, *Carbon Storage*, *Cmv* and *CO₂Value*) attending to the formulas explained before. These calculations were made directly on Power Bi data visualization tool.

In the same way the dataset provides variables such parishes, zone, age, place planted (i.e., if the tree is in a school, street, or park area), projection, state and local analysis, beyond the geographical coordinates, we could combine it with previous values calculated and the species trees. The data model was developed on Microsoft Power BI [see Appendix 4].

4.5. DATA VISUALIZATION

The motivation for the presented study was to get insights about the places, where trees are located, relating it with tree features and analysing their relationship with carbon storage and value. To fundament the study and give support to community decision-makers for trees plantation planning measures, we developed a reporting dashboard to respond to the following questions:

- In which zones of Viseu were higher values of carbon storage?
- What is the age of the trees? And how much *CO₂* is stored in those areas?
- Are the trees situated at schools, parks, or street? How many carbons can they store?
- What is the relation between the species tree and the carbon stored and value?

5. RESULTS

Here, we provide some experimental evaluation of the new technique. The approach adopted to analyse the study results include four perspectives: tree species, geographic analysis, age and state and type analysis. In each perspective we have in consideration the number of trees, carbon storage and CO_2 value.

5.1. TREE SPECIES

The most represented tree species was *Acer Negundo*, with 11.01% of individuals, followed by *Tilia Tomentosa* and *Liquidambar styraciflua* with 9.32%. and 7.51%, respectively. The presented dataset has a higher number of tree species and we cannot find a dominant species. Additionally, when we agglomerate the trees in *Others*, we create a most representative category species. The trees did not have missing values in its species and dimensions data. In *Figure 1* is presented the relative frequency for each tree species (Nh) from the total of 12,144 trees – after data preparation.

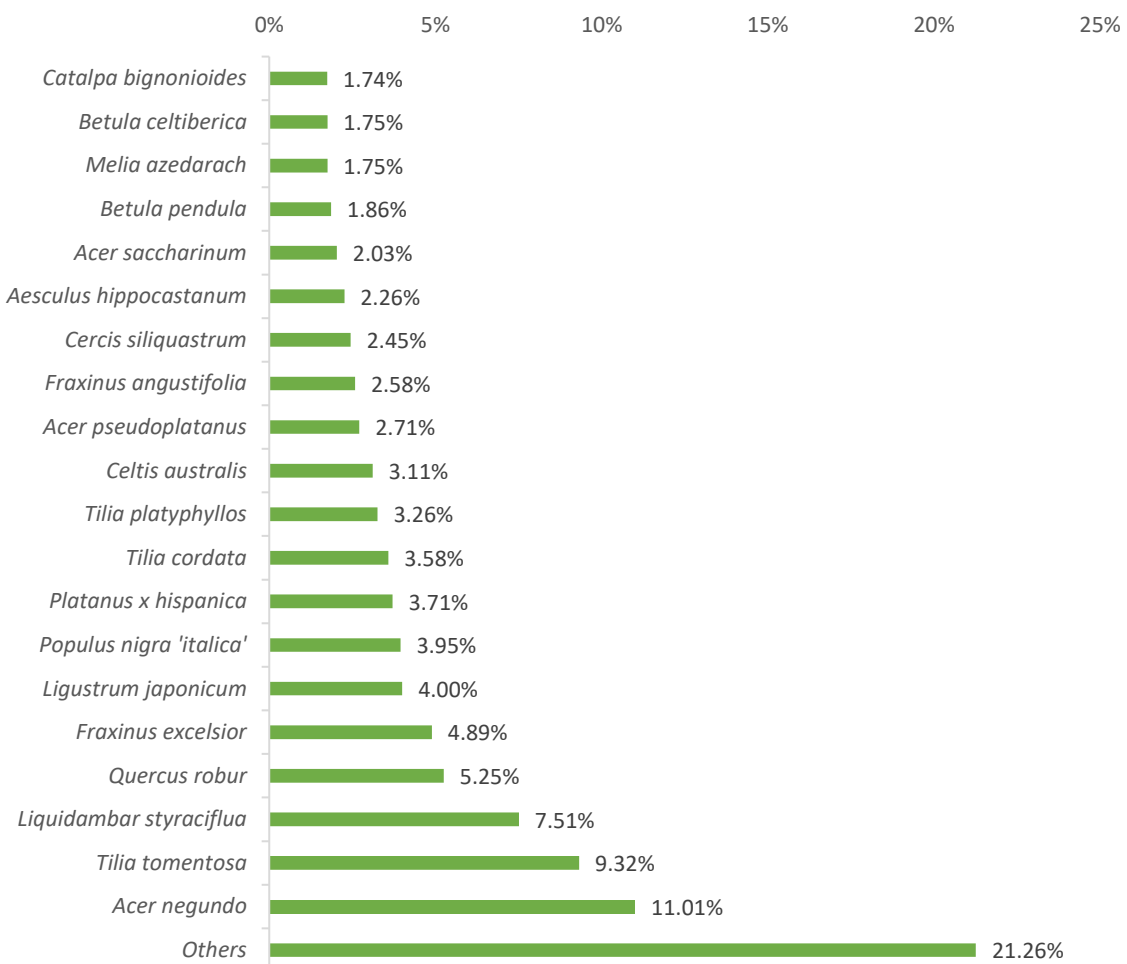


Figure 1: Trees relative frequency

The total carbon amount stored by the 12,144 trees of Viseu was 8,921Kg, with a higher contribution of the *Tilia tomentosa*, 2,572Kg and *Platanus x hispanica*, 2,548Kg. In the same way of the previous

analysis the Other Species represent a higher value, with 3,266Kg carbon stored, [see Appendix 5]. Therefore, we can observe that the most representative defined species do not represent the tree that absorbs more carbon, *Figure 2*.

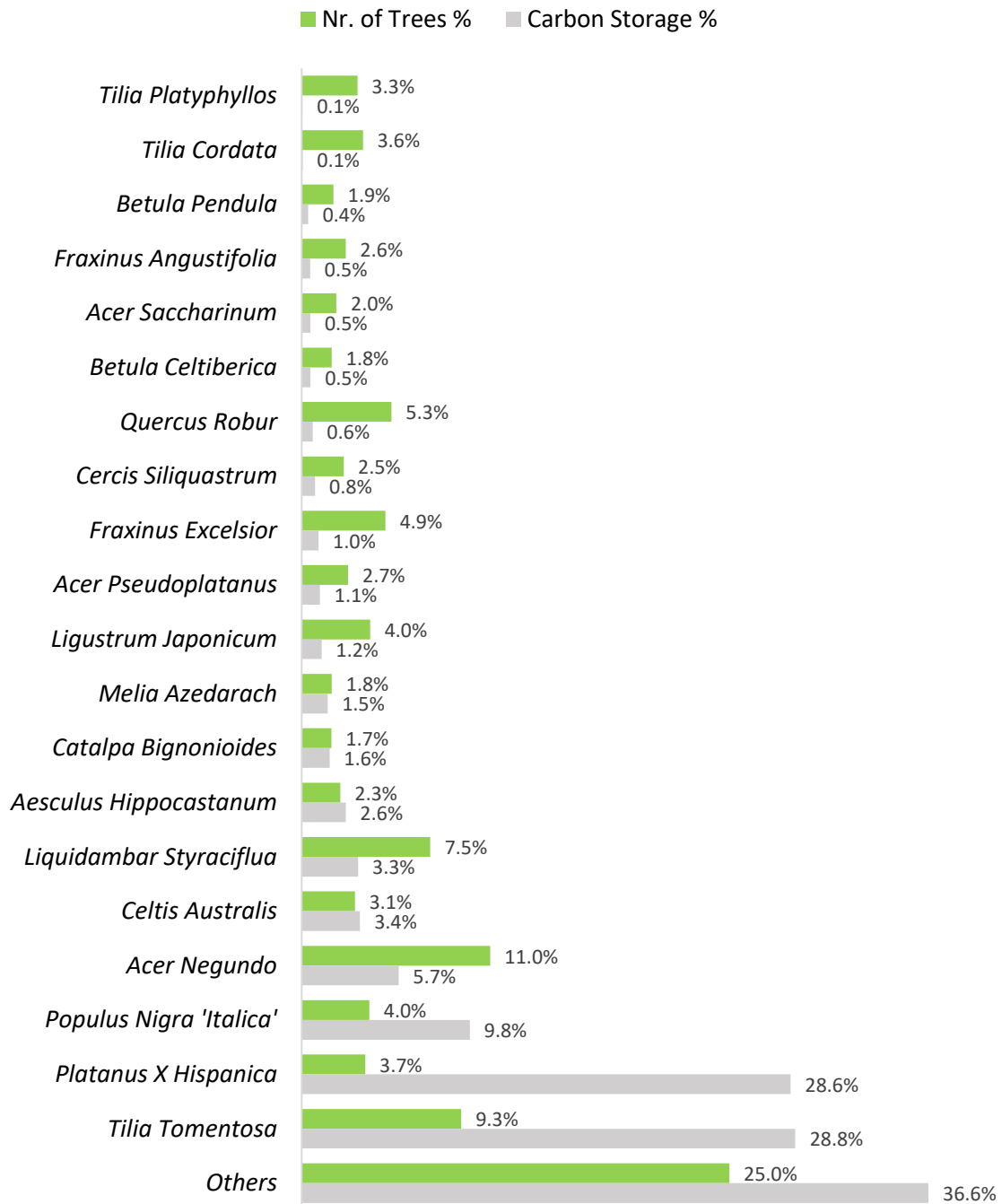


Figure 2: Number of trees and carbon storage, as a percentage, by species

At the same time, according to Figure 2 we can observe species that have higher C_n Value comparing to its number of trees. For example, *Melia azedarach* species with less than ½ of representative of *Quercus robur*, 213 and 637 trees respectively, have more than a double of the C_n Value: 135.63Kg and 56.99Kg [see ppendix 5].

According to the CO_2 value, as expected, we were able to detect the same behaviour by species. Agreeing to *Figure 3*, the species *Tilia tomentosa* is the one that gives the most credit for carbon dioxide to the city 32,801€, followed by *Platanus x Hispanica* 32,493€ and *Populus nigra 'italica'* 11,170€. For their part, the species included in the *Others* offer the largest credit € 41,649€. The species with the lowest stored carbon dioxide are those that, likewise, contribute less credit to Viseu.

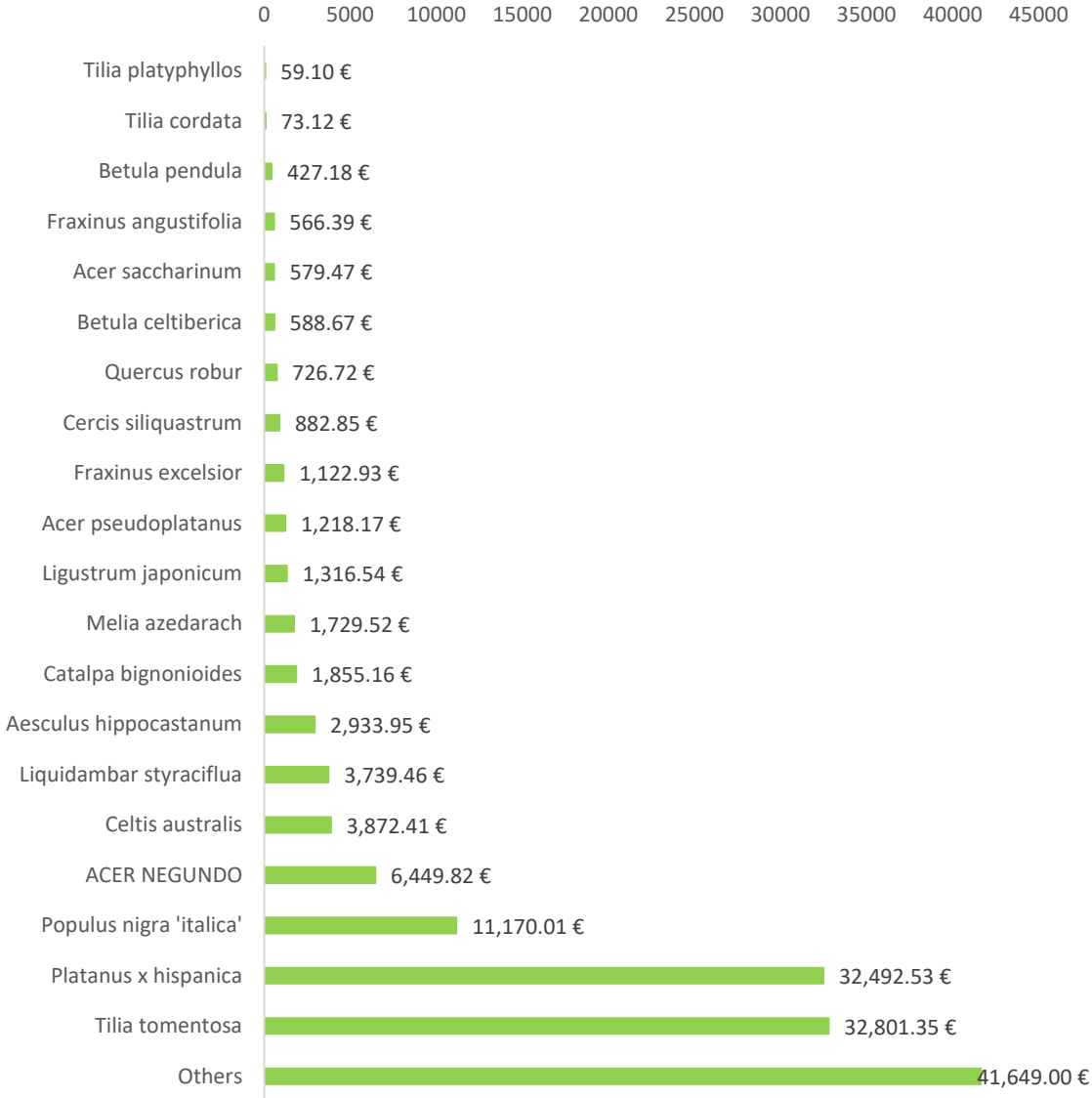


Figure 3: CO_2 Value per Specie

We found significant positive effect on carbon storage for tree species because the total carbon amount that was stored by species, like *Tilia Tomentosa*, *Platanus x hispanica* or *populus nigra 'italica'* are higher than *acer negundo* or *liquidambar*. This second species group represents more trees quantity when compared with first group mentioned.

At the same time, we found no significant consist effect for trees amount. So, the *figure 4* shows a positive relation between the CO_2 Value and number of trees. Although, the points scattered across the graph and the lower R2 indicate the security does not generally follow a linear behaviour.

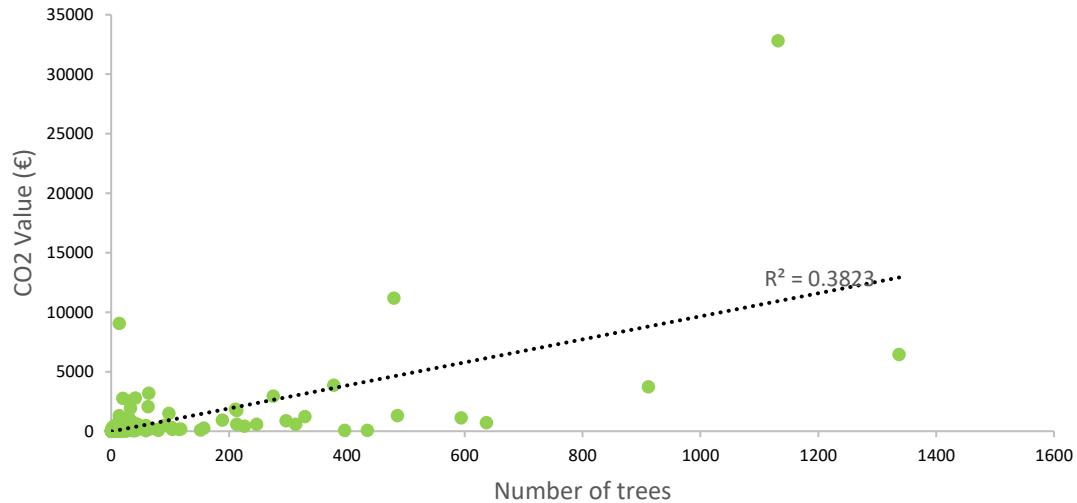


Figure 4: Number of trees and CO_2 relation

We applied different allometric equation to compute de trees volume so, we considered that finding common behaviour between the variables was unnatural. The higher values in dimension data could explain the higher CO_2 Value, however the same rationale could be associated to species like *Platanus x hispanica*, it has the higher tree branch, dbh_{cm} , h_m mean values.

5.1.1. Variables Correlation

The main goal before the analysis was done was to impute the outliers to have a segmentation with more reliable results. If the original outlier's value were included in this analysis, the results would change very significantly. As such, a boxplot for both variables, diameter at breast height e height, was used, allowing to see its distribution without doing any damage to the representativeness of the population [see appendix 6].

Consequently, per species, despite the positive relationship between the number of trees and CO_2 Value, we could not find an evident relation between them, *figure 5* and *figure 6*. If we had considered a linear relation between them, we would get a R square lower than 0.5 in both cases.

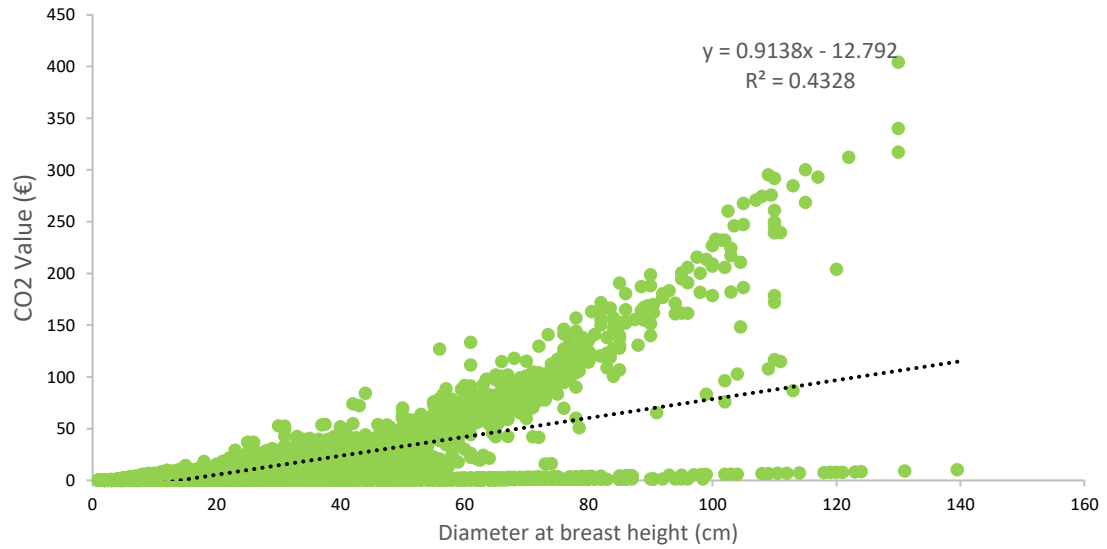


Figure 5: Diameter at breast height and CO_2 Value relation

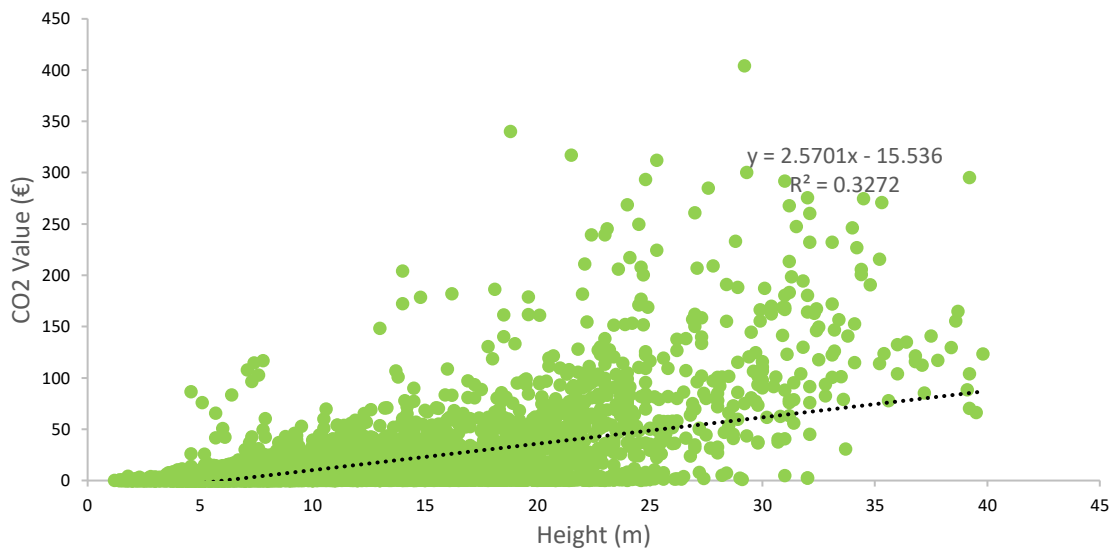
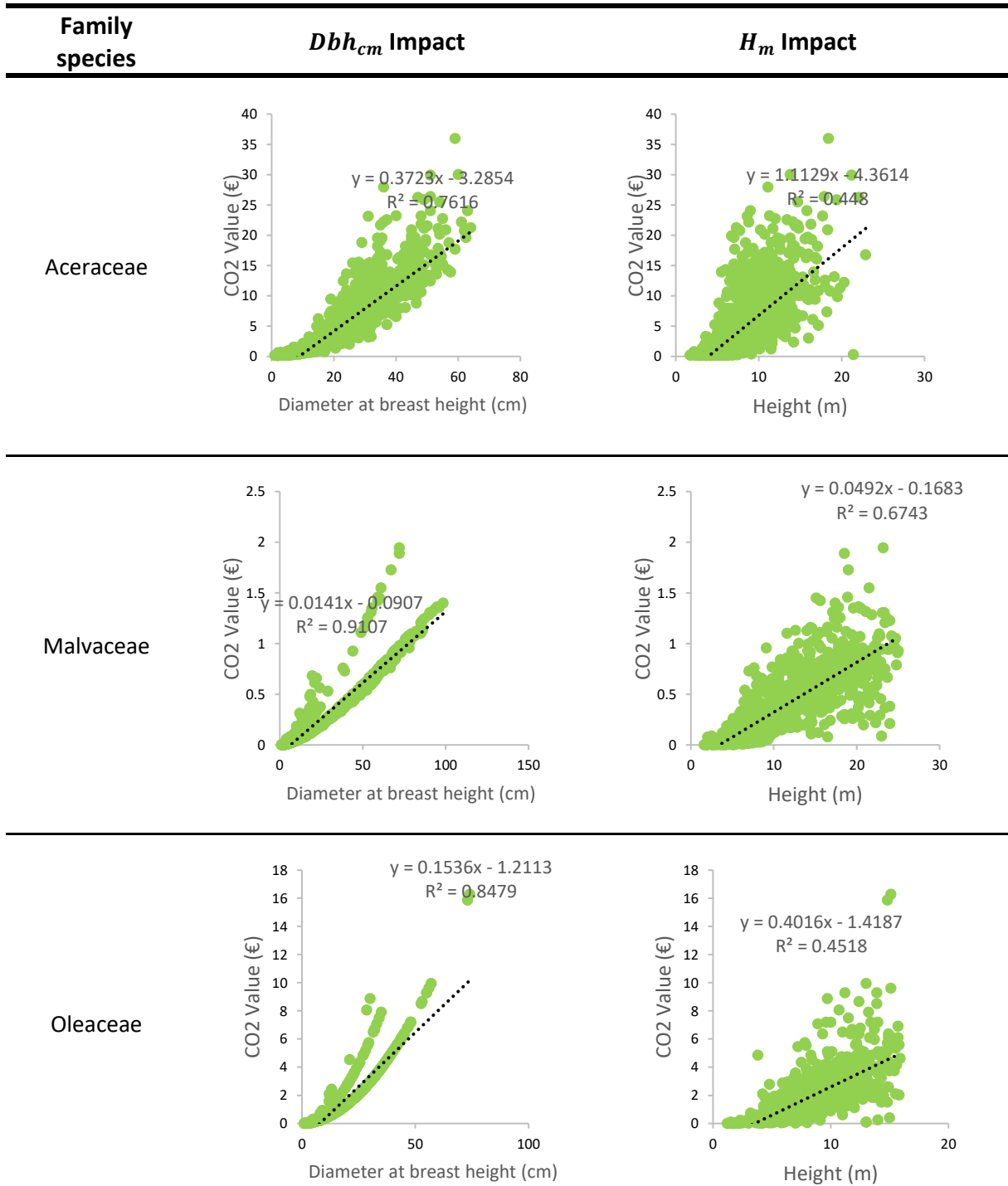


Figure 6: Height and CO_2 Value relation

One important issue in using the dbh_{cm} as a monitoring approach is to determine its impact on it. The impact is calculated by species family using the scatter plots of different composite CO_2 and dbh_{cm} , and CO_2 and h_m in each scene. A more detailed picture of the variability among estimates of credit CO_2 is presented in Table 7.

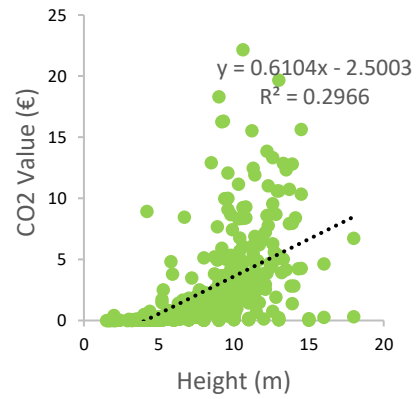
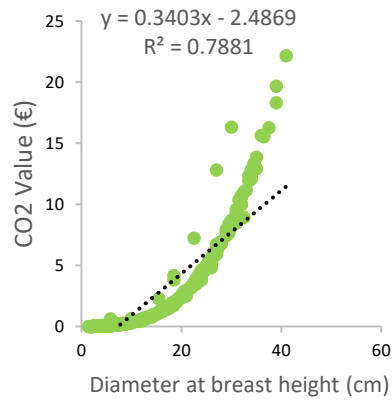


Family species

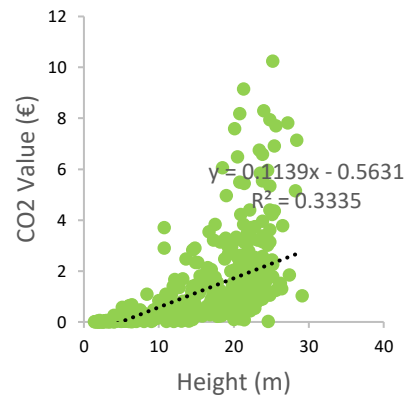
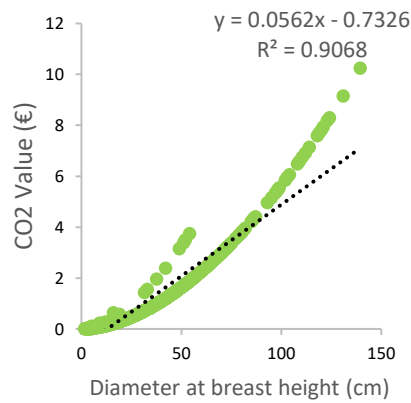
Dbh_{cm} Impact

H_m Impact

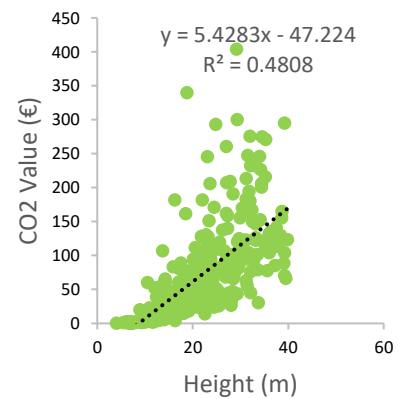
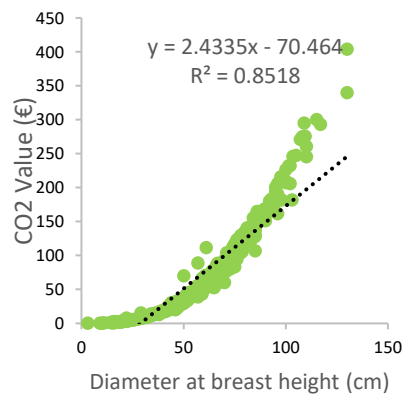
Betulaceae



Fagaceae



Platanaceae



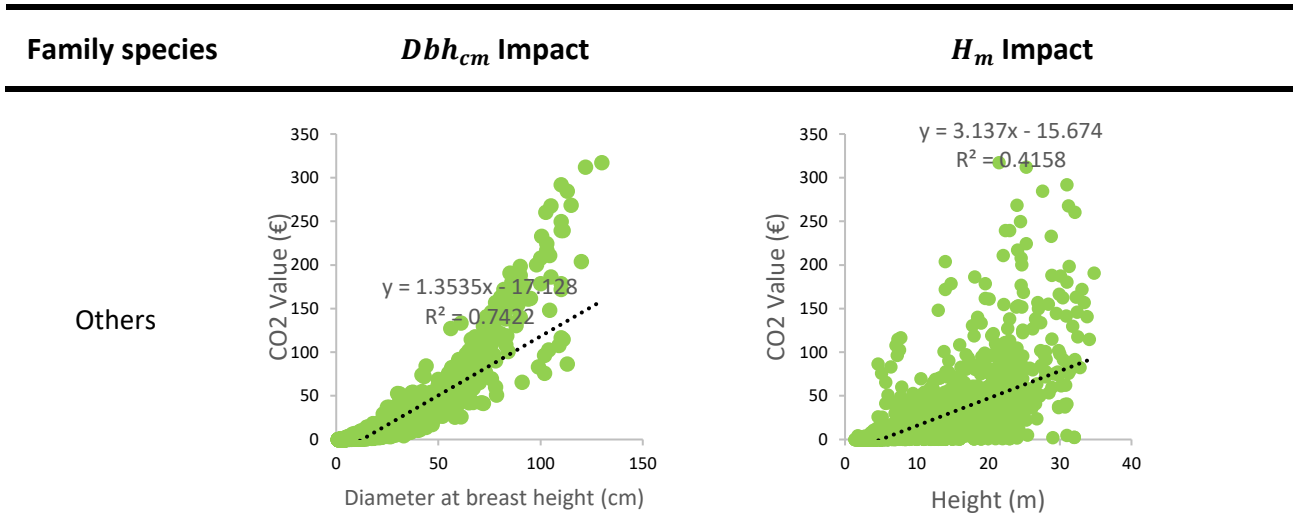


Table 7: Diameter at breast height and Height impact, by species family

Positive relation is observed for the both variables. When dbh_{cm} and h_m increase the credit of CO_2 value increases. These behaviours not necessarily look a linear relationship between both variables and CO_2 . Once the R-squared value is higher than 70% this contribute to our linear relations between the variables, because generally considered strong effect size. We can observe this aspect relatively to Dbh_{cm} to family *Aceraceae*, *Malvaceae*, *Betulaceae*, *Fagaceae*, *Platanaceae* family species. The *Malvaceae* present the higher R-square value for both variables. Attending to this family, a 10cm increase in dbh_{cm} will increase CO_2 by 0.050 €, Equally, a 10cm increase (10%) in h_m will increase the amount of CO_2 by (0,082)€The negative value, relative to CO_2 , may indicate the high oxygen consumption by this species. To better preview the variables impact on the CO_2 we had computed the equation derivatives [Appendix 7].

$$\Delta CO_2 = 0.0141(10dbh_{cm}) - 0.0907 = 0.0141(10) - 0.0907 \approx 0.050 \text{ €};$$

$$\Delta CO_2 = 0.0492(h_m + 10\%) - 0.1683 = 0.0492(1.6 \times 1.10) - 0.1683 \approx -0,082\text{€};$$

We can observe a higher difference between the variables impact. The presented values correspond to the expected values. Since trees have a higher number and volume of leaves in their branches, it is expected that the dbh_{cm} has a greater impact compared to h_m . Therefore, the higher dbh_{cm} values will increase the volume trees and the CO_2 credit.

To clarify reasons for different CO_2 amounts it is useful to examine differences among species that are most important by virtue of their abundance and size. For example, *Tilia Tomentosa* is not the most plentiful species, but it assumes the most CO_2 credit in accordance with all our equations.

There are no discernible trends in terms of a set of equations always producing C_{Kg} estimation that is the highest or lowest across all species. The dbh_{cm} and height impact will depend the equation used, but we can observe a positive impact in all of them. Attending to the species with more contribution to the CO_2 , like *Tilia Tomentosa*, *Fagaceae* *Quercus Spp* and *populus nigra 'italica'*, we can observe a relation closer to linear. So, the higher trees dimension values, the higher CO_2 credit.

5.2. GEOGRAPHIC ANALYSIS

Once the trees and respective values are concentrated on Viseu zone, we made the analysis at census tracts level (*Base Geográfica de Referência de Informação* – BGRI) – that includes the spatial representation of locals, sections and subsections. We represent the statistics subsection that represent the smallest territory entity for trees data provided by Viseu. The census area corresponds, at least, to the area where we can find one tree.

When it comes to smaller territories, the urban analysis requires updated data on the trees. For this portrayal and examination, registration information was used. This evaluation information is just helpful for urban arranging if they have an enormous spatial disaggregation. Along these lines, they should be gathered by urban territory units of a measurement that empowers huge scope urban arranging (registration tracts).

In *figure 7*, we represented the subsections corresponding to the city of Viseu. The darkest green represents higher carbon value provided by the trees, in contrast to the lower values. The silver values represent census blocks that do not have the presence of any tree. Bearing in mind that the zone with the highest representation of trees is the zone of Viseu, most of the area represented corresponds to this zone.

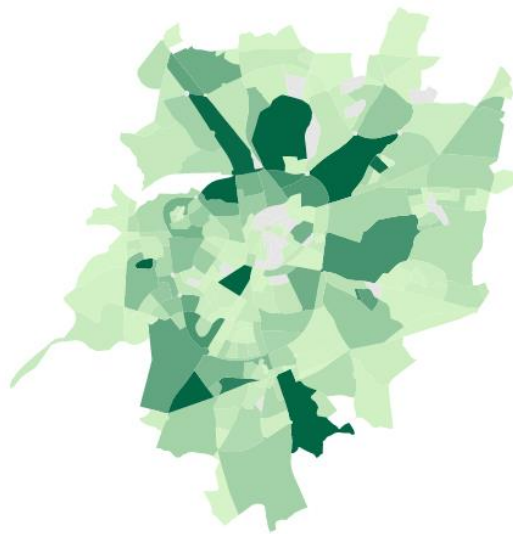


Figure 7: Carbon value in Viseu, by census tract

The number of trees presented in each zone have a significant difference among them. The zones that have the highest CO_2 Value are Viseu and Ranhados, as the storage carbon and number of urban trees. Viseu contribution to credit CO_2 Vaue, € 108,894.34, with 10,703 trees, and Ranhados with 1,130 trees gives € 3,943.96 credit. Low carbon monetary values were presented in the zones of Rio de Loba and Santa Maria, *table 8*.

| Locality | Nr. of trees | Carbon Stored (kg) | CO ₂ Value (€) |
|------------------------|--------------|--------------------|---------------------------|
| Viseu | 10,703 | 8539.67 | € 108,894.34 |
| Ranhados | 1,130 | 309.29 | € 3,943.96 |
| Repeses e São Salvador | 209 | 54.65 | € 696.84 |
| Coração de Jesus | 60 | 10.47 | € 133.55 |
| Rio de Loba | 31 | 7.21 | € 91.88 |
| Santa Maria | 11 | 0.10 | € 1.27 |
| Total | 12,144 | 8,921.39 | € 113,761.83 |

Table 8: Measures by locality

5.3. TYPE

Attending to the type of the trees, we need to clarify that this variable informs if a tree is located at park, school, or street *table 9*.

| Type | CO ₂ Value (€) | CO ₂ Value (%) |
|--------|---------------------------|---------------------------|
| School | € 3,014.45 | 2.65% |
| Park | € 25,023.53 | 22.00% |
| Street | € 85,723.85 | 74.35% |
| Total | € 113,761.83 | 100% |

Table 9: CO₂ value and relative frequency by location type.

Analysing *table 9*, the highest values of carbon storage are located on Street representing approximately 74% of the carbon value. At the same time, 22% of the carbon are stored on a Park in a and just 2.65% are stored in a school. Globally at schools, we just have 307 trees and it just stores €3,014.45 Carbon. The carbon monetary value is higher on streets, with € 85,723.85.

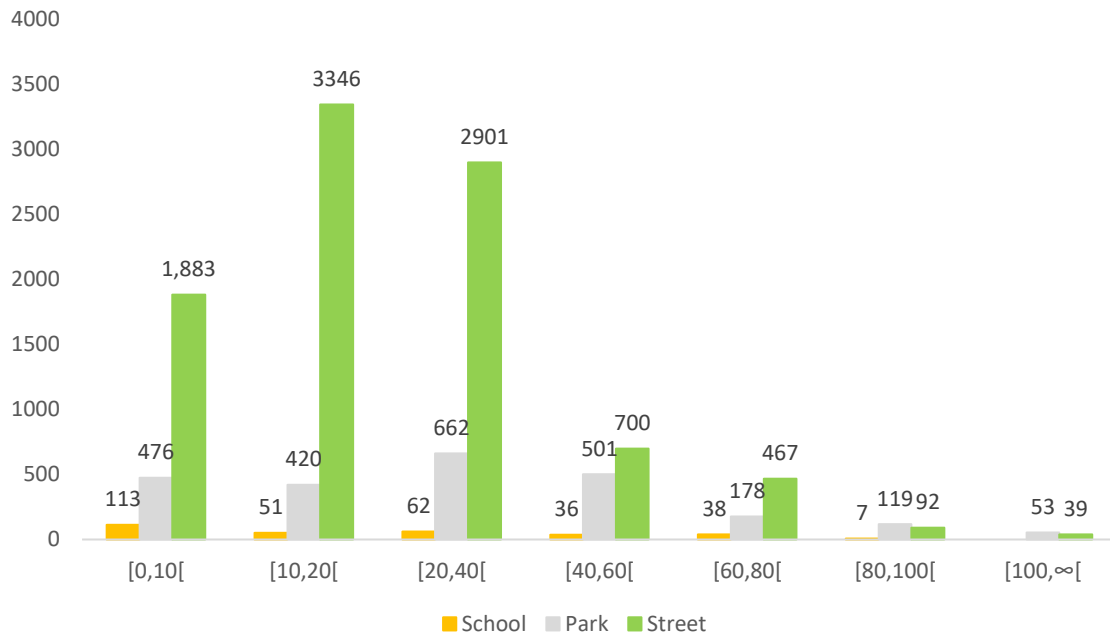


Figure 8: Number of trees by Type and Age.

Considering *figure 8*, over the most recent 10 years we have attempted to build the number of trees in schools, multiplying the worth. As far as concerns its, over the most recent couple of years, the quantity of trees planted on the avenues has diminished to half. Concerning trees planted in the parks, over the most recent 60 years, the normal increment per ten years is roughly 343 trees. The worldwide number of trees has developed.

5.4. AGE AND STATE

The break temp with higher CO_2 Value is the 20-40 years, following by the 60-80 years and 40-60 years, *figure 9*. In the past 20 years, planted trees have managed to offer around 7% of the total CO_2 credit. In view of the past 80 years, this period was the one with the least contribution.

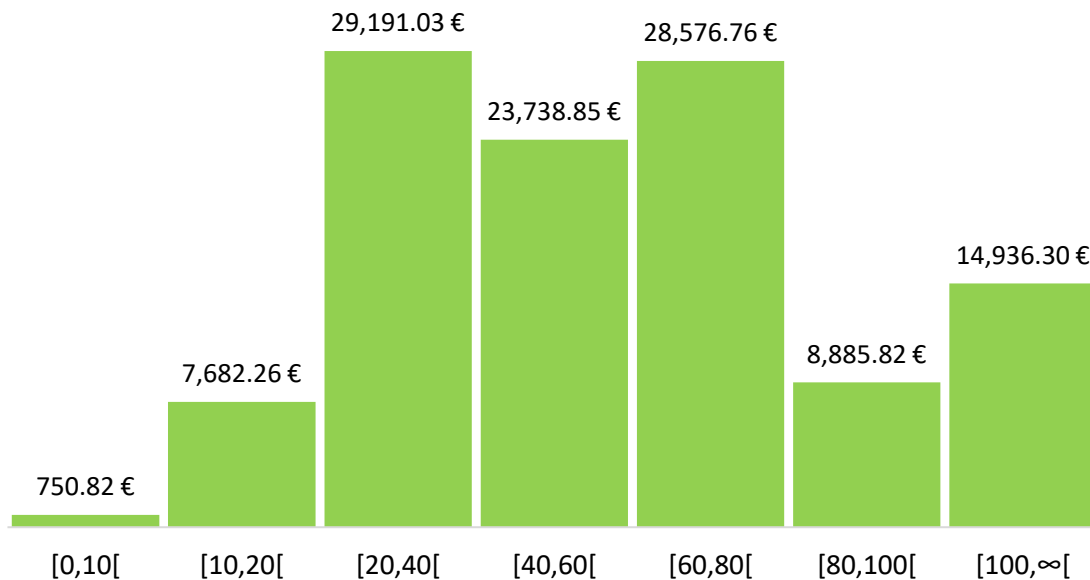


Figure 9: CO_2 value, by age

We can also relate the trees age and state, by number of trees. Emphasis on *figure10*, approximately 5% of the younger trees are weak, decrepited, or dead. At the same time, the more than 50% of the oldest trees are good and more than 60% of the trees data are in good or great state. Another important conclusion was that more than a half of the trees were planted 20-60 years ago, and in the last ten year, based on store data, just 20% of the trees were planted [see appendix 7].

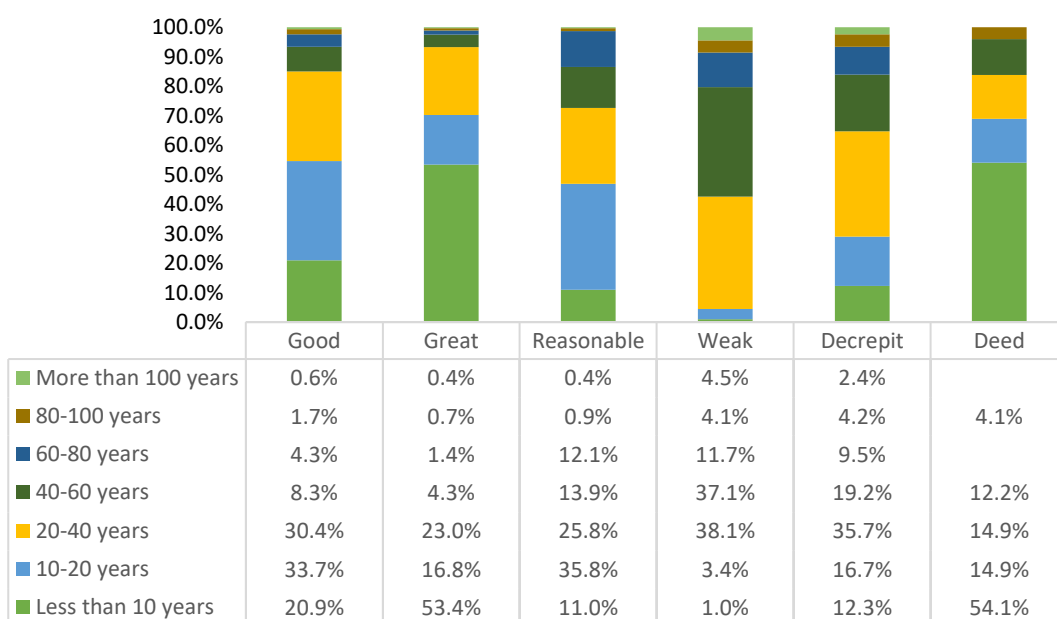


Figure 10: Age and state relation

5.5. REPORTING DASHBOARDS

The dashboard provided contains four pages to respond to the questions presented before. Each page has the tree measures: Nr. Of Trees, Carbon Storage (Kg) and CO_2 Value (€). As the three metrics are always present, on each sheet of the dashboard, it is possible for us to have access to their value during attribute comparisons and before different analyses. On the other hand, it is a common index for all the six pages, that gives the capability to the user to cross easily between pages. We provide the dashboard in Portuguese and English, and to change cross the language, we just need to click on the flags and the information provided in both opinion languages are the same.

5.5.1. Overview

The overview is the first dashboard page and shows the total values of the tree measures analysed: Nr. Of Trees, Carbon Storage (Kg) and CO_2 Value (€). Moreover, it presented the five key sections of the study: Species, details, geographic, variable relation, and information, *figure 11*.

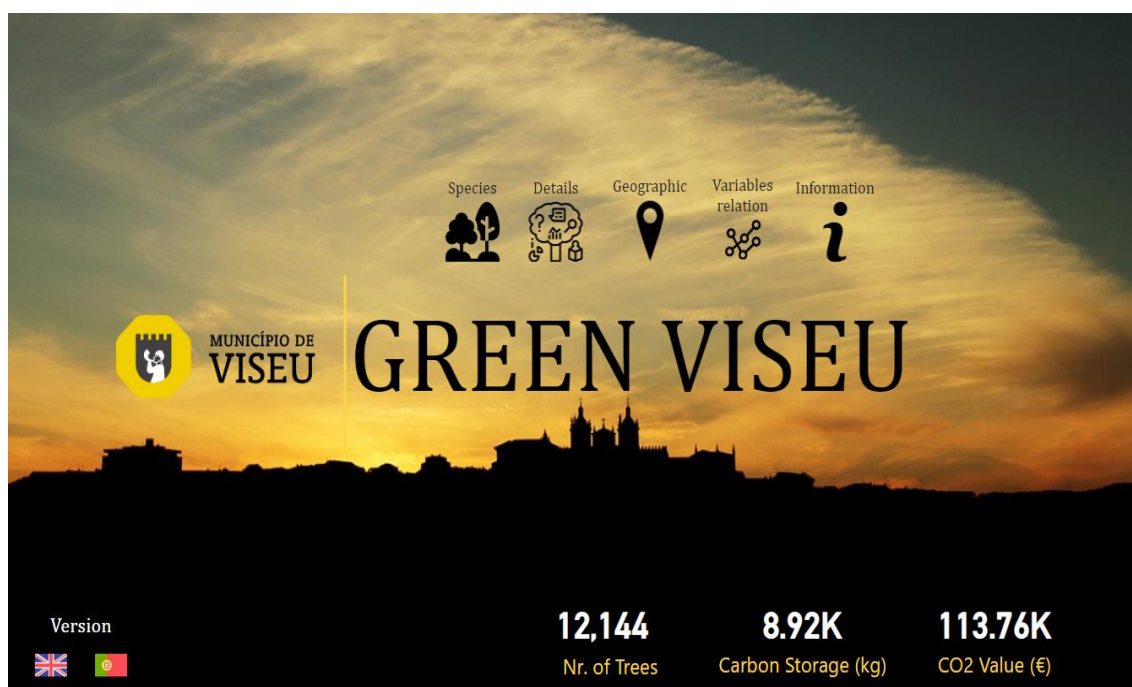


Figure 11: Overview report, english version

5.5.2. Species

The Home screen is composed by the main indicators on the species of trees in Viseu. In the middle we have the top 10 tree species, considering the available CO_2 credit. On the right side we have a graph where we can compare the number of trees and Carbon Storage according to the species. So, it

will be possible to analyse the relation between the species tree and the carbon stored and value. The *figure 12* shows the Species report.

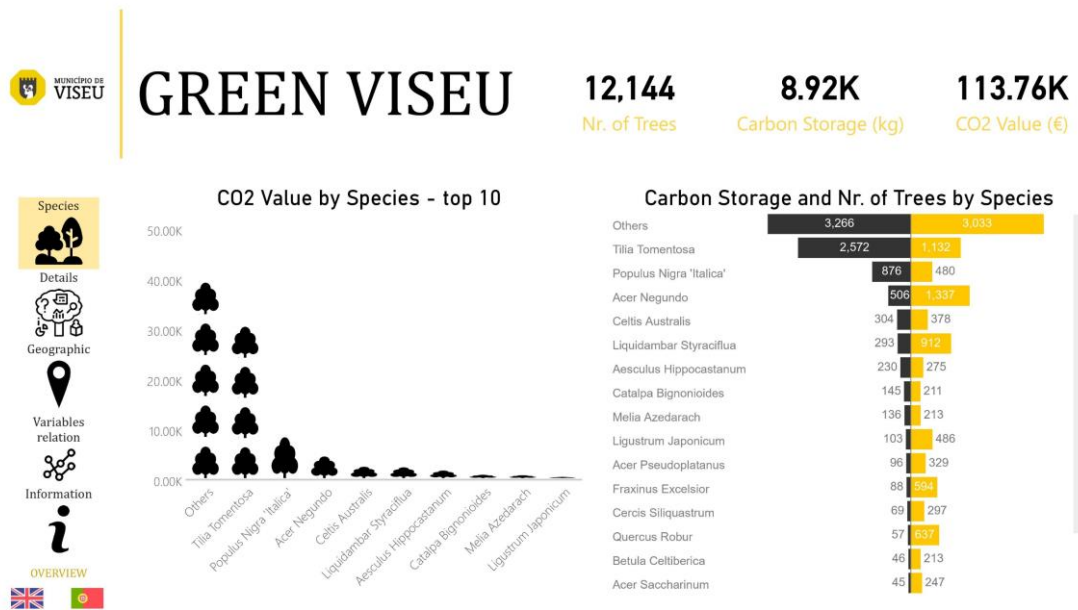


Figure 12: Species report, english version

5.5.3. Details

In the Details section, *figure 13*, we present more information about age, area, and type. This screen presents the aggregated information that can be filtered by the graph's attributes, in this way we were able to allow different combination attributes during the analysis, because of the interactivity of the dashboards. In addition, in the bottom of the dashboard, we have a table with information about the species, the zone, the address of its location and the state of the location where it is located, for each of the three metrics explored throughout the dashboard. We provide the user the possibility to consult, during their analysis of the graphs, more specific information about the trees, in the table, according to the attributes previously mentioned. Consequently, it becomes possible to understand the age of the trees, how much CO₂ is stored by the different areas (street, park or school), to understand the age of the trees in the streets, school and parks In addition, we could also add the details of the state of the trees. In other terms, we became possible to draw conclusions about the state of the trees in relation to their age and location.

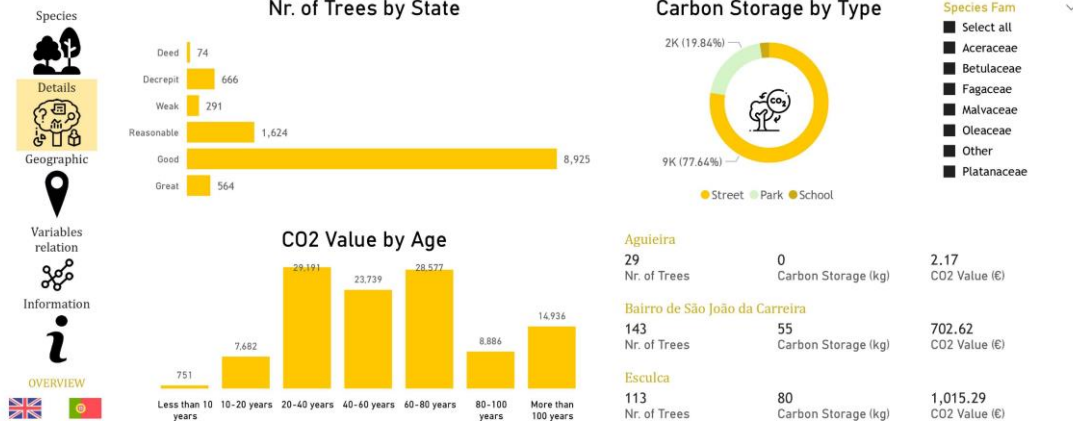


Figure 13:Details report, english version

5.5.4. Geographic

On the second page of the dashboard, *figure 14*, on the left side, we have a graph where it is possible to analyse which zones have more and less trees. On the opposite side, we have two maps of the city of Viseu. This allows us to detail carbon storage and by type and species. It is possible to combine the type of the trees with their geographic location and species. So, the present sheet will enable us to answer to which zones of Viseu are there higher values of trees, carbon storage, and CO_2 credit.

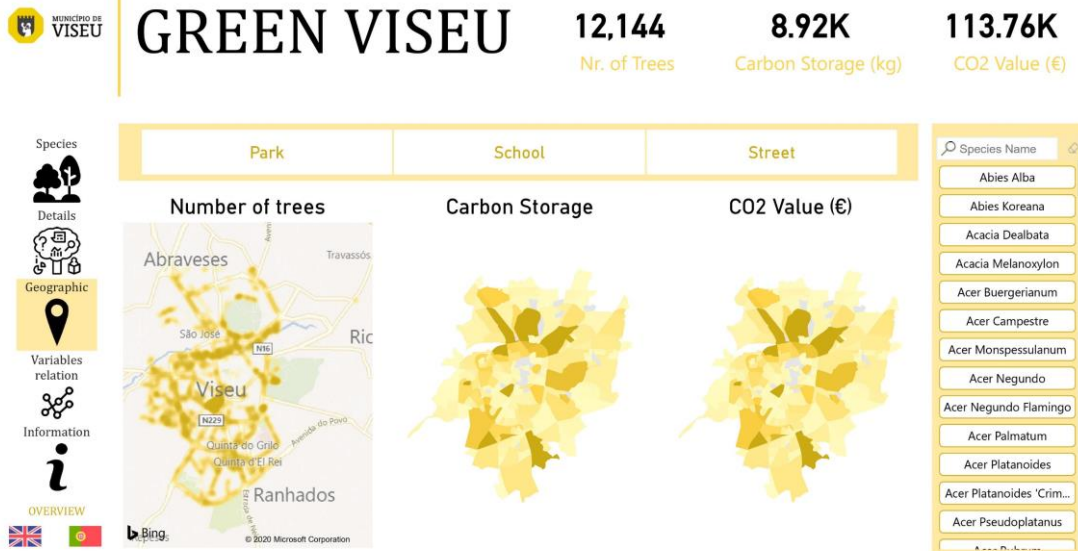


Figure 14: Geographic report, english version

5.5.5. Variables Relation

This page, *figure 15*, represents the preview variables relation analysis. We awarded the dependent variables, Dbh_{cm} and CO_2 , behaviour and impact on CO_2 credit value, evidencing the relation between both variables. We provide two scatter graphs, one for each dependent variable, and a filter by family species.

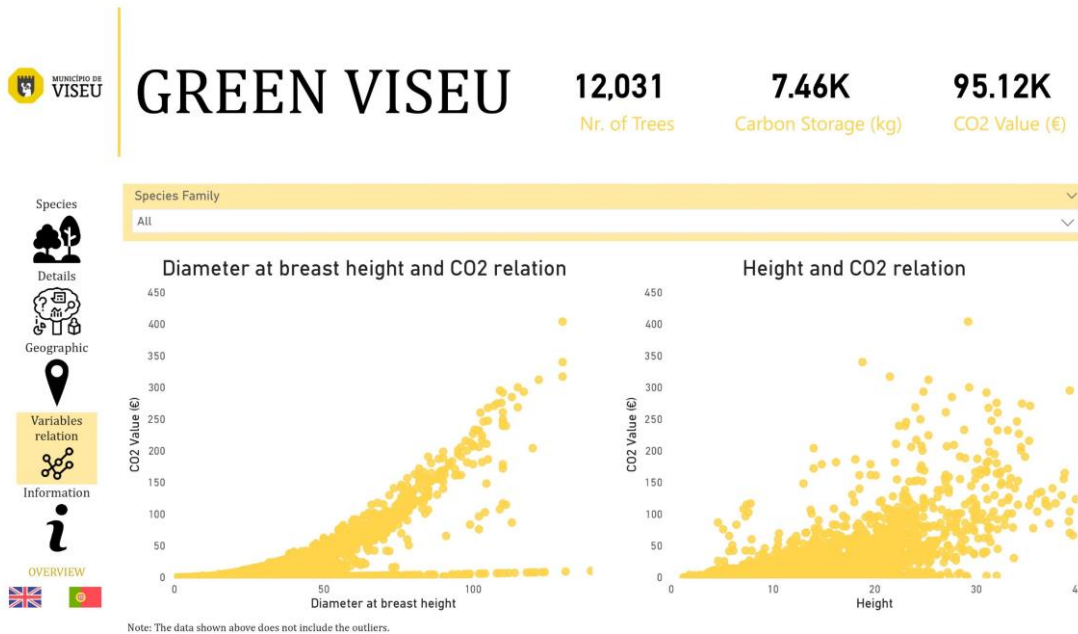


Figure 15: Geographic report, english version

5.5.6. Information

The last sheet is about Viseu city and gives to the user information about the city: geographic location, and project goals. We Describe the data and the purpose of the dashboard. Additionally, we clarify the measures conceived, *figure 16*.



Figure 16: Information report, english version

6. CONCLUSIONS

In this study, it was disclosed the carbon monetary value and storage based on the twenty most frequent tree species for the Viseu urban area. One important conclusion about the data was the diversity of the species, considering we found 148 different species. It is important to consider that the present study was only carried out by only considering the grouping of the 148 species into groups of the seven most represented tree species. In that sense, the use of specific allometric equations for each species of tree to compute the carbon storage can lead to different outcomes.

A total of 12,144 trees in Viseu capture 8,921.39Kg of CO_2 , which corresponds to €113,761.83 in terms of CO_2 value. The *Acer Negundo* species was the most common species in the sample with 1,337 trees and storing 505.81Kg of carbon, although the *Tilia Tomentosa* was the species that store more carbon, with 2,572.34Kg. Furthermore, the majority of the trees are in good conditions but 5% of them with less than ten years old are weak, decrypted, or Dead. In addition, more than 6 in each 10 trees with more than 100 years are in good or great conditions.

The data collected was concentrated mainly in the parish of Viseu. The carbon storage presents higher values on streets €85,723.85 which corresponds to 6,722.61kg of storage carbon. So, the higher percentage of credit carbon offered by trees is based on streets. The lesser value is at schools, just 2.65% of both values. The parks show an intermediate value, but it only contributes with just 22%. Although we do not have full information about the trees in the Viseu city, there is a great disparity about the environments where the trees are located.

Despite working with open data, some categorical variables are missing, and we found incoherent values, which is an important drawback in the making of a more detailed future study. We removed around 35 of the observations by default values and imputed 6,811 of missing values, using decision trees based on available data. Parallely, we did not find allometric equations of all the 148 tree species so, we adopted seven tree species allometric equations that share similar Mediterranean climate or geographical conditions to compute the tree biomass and fundament the carbon value and storage calculated. This fact always precedes to a certain amount of error in the volume computation of tree species and, as a result, in the tree biomass and contribution to the storage Carbon calculation in this study could be biased. One more important feature, that might have influenced the result was the dry weight density factor used, as constant, which was not adapted to the geographic characteristics of the location.

Ultimately, to support the decision making on urban green infrastructures management, the present study includes an interactive dashboard, developed on Power BI. In addition to respond to the study goals, it should allow to combine different variables and provide an interactive approach of the data.

6.1. LIMITATIONS AND RECOMMENDATIONS FOR FUTURE WORKS

The Viseu City has an area of 507.10 km², with 24 parishes. In the presented study, we only collected data of six parishes, where we did not have completed information about all trees in the parishes included. So, we cannot estimate how much these trees mean in the total trees of Viseu. As we already referred, the data quality and some adopted assumptions could influence the results and, consequently, conclusions could possibly be bias and incorrect.

Additionally, along with the study, we had into consideration different approaches and viewpoints to increase the study value and the precision of decision making. Hence, we have in mind the completion of the study with a simulation of an increase of determined species.

7. BIBLIOGRAPHY

- Albino, V., Dangelico, R. M. & Berardi, U., 2015. Smart Cities: Definitions, Performance, and Initiatives. *Journal of Urban Tecnology*, February.
- Andersson-Sköld, Y. et al., 2015. An integrated method for assessing climate-related risks and adaptation alternatives in urban areas. In: *Climate Risk Management*. s.l.:s.n.
- Anon., 2011. Data Mining: Practical Machine Learning Tools and Techniques. In: s.l.:s.n.
- Anon., n.d. *CELL SITE ANALYSIS from CDRs*. [Online] Available at: <https://www.cellularforensics.com/cell-site-tower-analysis.html>
- CBO, 2012. *Cities and Biodiversity Outlook – Action and Policy*, s.l.: s.n.
- CemilBilgili, B. & Gökyer, E., 2012. *Urban Green Space System Planning, Landscape Planning*. s.l.:s.n.
- CIWEM, 2010. *Multi-Functional Urban Green Infrastructure*, s.l.: s.n.
- Climate change and Urban Vulnerability in Africa (CLUVA), 2013. *Green Infrastructure: An essential foundation for sustainable urban futures in Africa*, s.l.: s.n.
- College of Agriculture & Life Sciences, 2019. *Americans are planting . . . TREES OF STRENGTH*. [Online] Available at: <https://projects.ncsu.edu/project/treesofstrength/benefits.htm>
- Correia, I. P., 2012. *Das Estruturas Ecológicas Municipais*, Lisbon: s.n.
- Domingo, A. et al., 2013. Public open sensor data: Revolutionizing smart cities. In: *IEEE Technology and Society Magazine*. s.l.:IEEE, pp. 50-56.
- ECOTEC, 2008. *The economic benefits of green infrastructure: the public and business case for investing in green infrastructure and a review of the underpinning evidence*, s.l.: Natural Economy Northwest (NENW).
- Environmental Defense fund, n.d. *The true cost of carbon pollution*. [Online] Available at: <https://www.edf.org/true-cost-carbon-pollution> [Accessed 04 02 2020].
- Fadigas, L., 2010. *Urbanismo e Natureza - Os desafios*. Lisbon: Edições Sílabo.
- Fayyad, U., 2010. Knowledge discovery in databases: An overview. In: *Lecture Notes in Computer Science*. s.l.:s.n., pp. 5-16.
- Filippova, O. & Rehm, M., 2014. Cell phone towers and house prices in New Zealand: economic effects and policy implications. *International Journal of Housing Markets and Analysis*, pp. 18-29.
- Forest Research, 2010. *Benefits of green infrastructure*, Farnham: Report to Defra and CLG.
- Gaston, K. J., Ávila-Jiménez, M. L., Edmondson, J. L. & Jones, J., 2013. Review: managing urban ecosystems for goods and services. *Journal of Applied Ecology*, pp. 830-840.

Granitto, P. M., Gasperi, F., Trainotti, E. & Furlanello, C., 2007. Modern data mining tools in descriptive sensory analysis: A case study with a Random forest approach. In: *Food Quality and Preference*. s.l.:s.n., p. 681-689.

Greater London Authority (GLA), 2008. *The London Climate Change Adaptation Strategy*, London: s.n.

Hammer, S. e. a., 2011. *Cities and Green Growth: A Conceptual Framework*. [Online] Available at: <http://dx.doi.org/10.1787/5kg0tflmzx34-en> [Accessed 09 2019].

Hand, d., 2013. *Data Mining Based in part on the article "Data mining" by David Hand, which appeared in the Encyclopedia of Environmetrics*. [Online] Available at: <https://onlinelibrary.wiley.com/doi/abs/10.1002/9780470057339.vad002.pub2>

Hartig, T., Davis, D. S., Jamner, L. D. & Jamner, L. D., 2003. Tracking restoration in natural and urban field setting. *Journal of Environmental Psychology*, p. 109–123.

Heinze, J., 2011. *Benefits of Green Space – Recent Research*, s.l.: s.n.

Huang, M., Fang , Z., Xiong, S. & Zhang, T., n.d. Interest-Driven Outdoor Advertising Display Location Selection Using Mobile Phone Data. *IEEE*.

Husch, B., Beers, T. W. & Kershaw, J. A., 2003. *Forest mensuration*. : . 443 p.. 5 ed. New york (NY): John Wiley and Sons.

Ilie-zudor, E., 2014. *Journals & Books - Advanced predictive-analysis-based decision support for collaborative logistics networks*. [Online] Available at: <https://www.emeraldinsight.com/doi/full/10.1108/SCM-10-2014-0323>

Jones, A., M., H. & Coombes, E., 2009. Greenspace access, use, and physical activity: Understanding the effects of area deprivation, , 49(6), 500–505. In: *Preventive Medicine*. s.l.:s.n.

Kumar, A. S., Raja, B., Pamina, J. & Madhavan, P., 6. *Market Behavior Analysis using Descriptive Approach*. [Online] Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3330017

Kuo, F., 2001. Coping with poverty: impacts of environment and attention in the inner city. In: *Environment & Behaviour*. s.l.:s.n., pp. 5-34.

Lavalle, S. et al., 2011. Big Data, Analytics and the Path From Insights to Value. *MIT Sloan Management Review*, pp. 21-32.

Leith, H., 1963. The role of vegetation in the carbon dioxide content of the atmosphere. *Journal of Geophysical Research*, p. 3887–3898.

Máchová , R. & Lnénicka, M., 2017. Evaluating the Quality of Open Data Portals on the National Level. *Journal of theoretical and applied electronic commerce research*.

Mark Scott, M. L. D. H. A. K. G. C. & T. B., 2016. Nature-based solutions for the contemporary city. *PLANNING THEORY & PRACTICE*, Volume 17, pp. 267-300.

- McHale, M. B. I. L. M. e. a., 2009. *Urban Ecosyst.* 21 01.
- McPherson, E. A. a. E. G., 2012. Comparison of Methods for Estimating Carbon. In: *Urban Ecosystems and Social Dynamics Program*. s.l.:s.n., pp. 45-50.
- Mcpherson, E. G. & Doorn, N. S. V., 2016. *Urban Tree Database and Allometric Equations*, New York: United States Department of Agriculture.
- Mohankumar, D., 2016. *Mobile Phone Communication. How it works?*. [Online] Available at: <https://www.electroschematics.com/5231/mobile-phone-how-it-works/>
- NASA, 2010. *Global Warming*. [Online] Available at: <https://earthobservatory.nasa.gov/features/GlobalWarming>
- NASA, 2018. *Global Climate Change*. [Online] Available at: <https://climate.nasa.gov/evidence/>
- Neto, M. C., 2018. *Urban intelligence*, Santarém, Portugal: In Proceedings of the 18a Conferência da Associação Portuguesa de Sistemas de Informação (CAPSI 2018).
- Nowak, D. J. & Crane, D. E., 2002. Carbon storage and sequestration by urban trees in the USA. In: *Environmental Pollution*. s.l.:s.n., pp. 381-389.
- Ojo, A., Curry, E. & Zeleti, F. A., 2015. *A tale of open data innovations in five smart cities*, Kauai, HI, USA: s.n.
- Okunlola, A., 2013. Sustainable Environmental Management through Urban Public Park in Southwestern Nigeria. *Journal of Agricultural Science* .
- OMTEL, 2018. *Sobre*. [Online] Available at: <https://www.omtel.pt/pt/#hero>
- Rafferty, J. P., 2019. *Biodiversity loss*, s.l.: Encyclopaedia Britannica.
- Rahman, G. & Zahidul, I., 2013. Missing Value Imputation Using Decision Trees and DecisionForests by Splitting and Merging Records: Two NovelTechniques. 06 08.
- Renewable Energy Policy Network for the 21st Century, 2019. *Renewables in cities: 2019 Global status report*, s.l.: s.n.
- Ren, Z. & Zhai, C., 2019. Changes in spatio-temporal patterns of urban forest and its above-ground carbon storage: Implication for urban CO2 emissions mitigation under China's rapid urban expansion and greening. In: *Environment International*. s.l.:s.n.
- Rita, P., 2019. *Secondary Data*. Lisbon: s.n.
- Ruijter, E., Grimmelikhuijsen, S. & Meijer, A., 2017. Open data for democracy: Developing a theoretical framework for open data use. In: *Government Information Quarterly*. s.l.:s.n., pp. 45-52.
- Samora-Avela, A. et al., 2017. Green infraSTRUctUre, climaTe chanGe and SPaTial Planning: learning leSSonS acroSS BorDerS. *Journal of Spatial and Organizational Dynamics*, pp. 176-188.

Smyth, P., Piatetsky-Shapiro, G. & Fayyad, U., 1996. From Data Mining to Knowledge Discovery in Databases. *AI Magazine - Volume 17 Number 3*, pp. 37-54.

Soomro, T. R. & Sarwar, M., 2012. Impact of Smartphone's on Society. *European Journal of Scientific Research*, pp. 216-226.

Stempski, K. J. a. W., 2018. *An attempt to assess the monetary value*, Poznań, Poland: Series A – Forestry.

Stocker, T. F. et al., n.d. Summary for Policymakers. In: *Climate Change 2013: The Physical Science Basis*. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA: Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change.

Sussams, L. W., Sheate, W. R. & Eales, R. P., 2014. Green Infrastructure as climate change adaptation policy intervention: Muddying the waters or clearing a path to a more secure future?. *Journal of Environmental Management*, pp. 184-193.

The Nature Conservancy, 2016. *GLOBAL INSIGHTS | PERSPECTIVES: How Urban Trees Can Save Lives*. [Online]

Available at: <https://global.nature.org/content/healthyair>

The Sustainable Development Council (SDC), 2010. *Green Infrastructure*, s.l.: Comhar SDC.

Turcu, T., 2013. Re-thinking Sustainability Indicators: Local Perspectives of Urban Sustainability. *Journal of Environmental Planning and Management*, pp. 695-719.

Turnbull, J. C. et al., 2018. Synthesis of Urban CO₂ Emission Estimates from Multiple Methods from the Indianapolis Flux Project (INFLUX). *Environmental Science Technology*, 6 12.

United Nations , 2019. *Goal 13: Take urgent action to combat climate change and its impacts*. [Online]

Available at: <https://www.un.org/sustainabledevelopment/climate-change/>

United Nations, 1987. *Our Common Future*, s.l.: World Commission on Environment and Development.

United Nations, 2016. *Benefits of urban trees*. [Online]

Available at: <http://www.fao.org/resources/infographics/infographics-details/en/c/411348/>

United Nations, 2018. *News, Home*. [Online]

Available at: <https://www.un.org/development/desa/en/news/population/2018-revision-of-world-urbanization-prospects.html>

Wall, S. et al., 2019. Real-world cell phone radiofrequency electromagnetic field exposures. In: *Environmental Research*. s.l.:s.n., pp. 581-592.

Wang, Y. et al., 2019. Analysis of wind farm output characteristics based on descriptive statistical analysis and envelope domain. In: *Energy*. s.l.:s.n., pp. 580-591.

Wigren, T., 2018. *Department of Information Technology*. [Online]

Available at: <http://www.it.uu.se/katalog/tw/research/LocationTechnology>

Wirth, R., 2000. *CRISP-DM: Towards a standard process model for data mining*. s.l.:Proceedings of the Fourth International Conference on the Practical Application of Knowledge Discovery and Data Mining.

World Bank Group, 2019. *State and Trends 2019*, Washington DC: s.n.

Yahya, S. I., 2019. The Use of Camouflaged Cell Phone Towers for a Quality Urban Environment - Koya City as a Case Study. *UKH Jurnal of Science and Engineering*.

Z., T. P., 2013. Environmental challenges of urbanization: A case study for open green space management.. *Journal of Agricultural and Environmental Management*.

8. APPENDIX

8.1. EXPLORING DATA

```
1 """
2 EXPLORING DATA
3 """
4 # IN[3]
5 # Basic Information
6 print(VISEU.shape)
7 print(VISEU.columns)
8 print(VISEU.info())
9 print(VISEU.count())
10 print(VISEU.head())
11 print(VISEU.tail())
```

Figure 17: Exploring Data Python Code

```
dtype='object')
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 12579 entries, 0 to 12578
Data columns (total 61 columns):
OBJECTID      12579 non-null int64
Tipo          12579 non-null object
id_rua        12579 non-null int64
Id_rua_arv    12579 non-null int64
designacao     12579 non-null object
freguesia     12579 non-null object
localidade    12579 non-null object
toponimo      12579 non-null object
data_fonte    12579 non-null object
data_concl    12579 non-null object
Cod_esp       12579 non-null int64
nome_cien     12579 non-null object
nome_comum    12579 non-null object
DAP_cm        12579 non-null float64
Altura m      12579 non-null float64
```

Figure 18: Part of Python Script results about Exploring Data

8.2. DATA COHERENCE

```
1 """
2 DATA COHERENCE
3 """
4 len(VISEU) != len(set(VISEU))
5
6 def del_with_cond (VISEU,cond1):
7     """
8     input: dataframe, condition
9     output:VISEU without rows after the application of condition
10    Function to delete certain rows with a condition
11    """
12    VISEU['boolean']=cond1
13    VISEU=VISEU.query('boolean==False')
14    VISEU.__delitem__('boolean')
15    return VISEU
16
17 # Summary Statistics
18 print(VISEU.describe().T)
19 VISEU_Statistics = VISEU.describe().T
20
21 # Summary Statistics
22 print(VISEU.describe().T)
23 VISEU_Statistics = VISEU.describe().T
24
25 VISEU_clean = VISEU.copy()
26
27 VISEU_1 = VISEU_clean[VISEU_clean['DAP_cm'] != 0]
28 VISEU_1 = VISEU_clean[VISEU_clean['Altura_m'] != 0]
29 VISEU_1 = VISEU_clean[VISEU_clean['Alt_ramo_m'] != 0]
30 VISEU_1 = VISEU_clean[VISEU_clean['d_copa_cm'] != 0]
31
32
33 print(VISEU_1.count())
34
```

Figure 19: Data coherence

8.3. HANDLING MISSING VALUES

```
1 ##Replace the missing value##
2 VISEU_2 = VISEU_1.copy()

1 import pandas as pd
2 import numpy as np
3 # explicitly require this experimental feature
4 from sklearn.experimental import enable_iterative_imputer # noqa
5 # now you can import normally from sklearn.impute
6 from sklearn.impute import IterativeImputer
7 from sklearn.ensemble import ExtraTreesRegressor
8 #from sklearn.neighbors import KNeighborsRegressor
9 from sklearn.preprocessing import OrdinalEncoder

1 import seaborn as sns
2 impute_data = VISEU_2.copy()
3 pd.options.display.max_columns = None
4 impute_data.head()
```

Figure 20: Handling missing values - python code I

```
1 #instantiate both packages to use
2 encoder = OrdinalEncoder()
3 imputer = IterativeImputer(ExtraTreesRegressor())
4 # create a list of categorical columns to iterate over
5 VCOLUMNS = ['Tipo', 'designacao', 'Idade', 'Envolvenc', 'Projecao_c', 'Anal_Local', 'Estado', 'Arv_intere', 'toponimo', 'nome_cien', '
6
7 def encode(data):
8     '''function to encode non-null data and replace it in the original data'''
9     #retains only non-null values
10    nonulls = np.array(data.dropna())
11    #reshapes the data for encoding
12    impute_reshape = nonulls.reshape(-1,1)
13    #encode date
14    impute_ordinal = encoder.fit_transform(impute_reshape)
15    #Assign back encoded values to non-null values
16    data.loc[data.notnull()] = np.squeeze(impute_ordinal)
17    return data
18
19 #create a for loop to iterate through each column in the data
20 for columns in VCOLUMNS:
21    encode(impute_data[columns])
```

Figure 21: Handling missing values - python code II

8.4. DATA MODELER

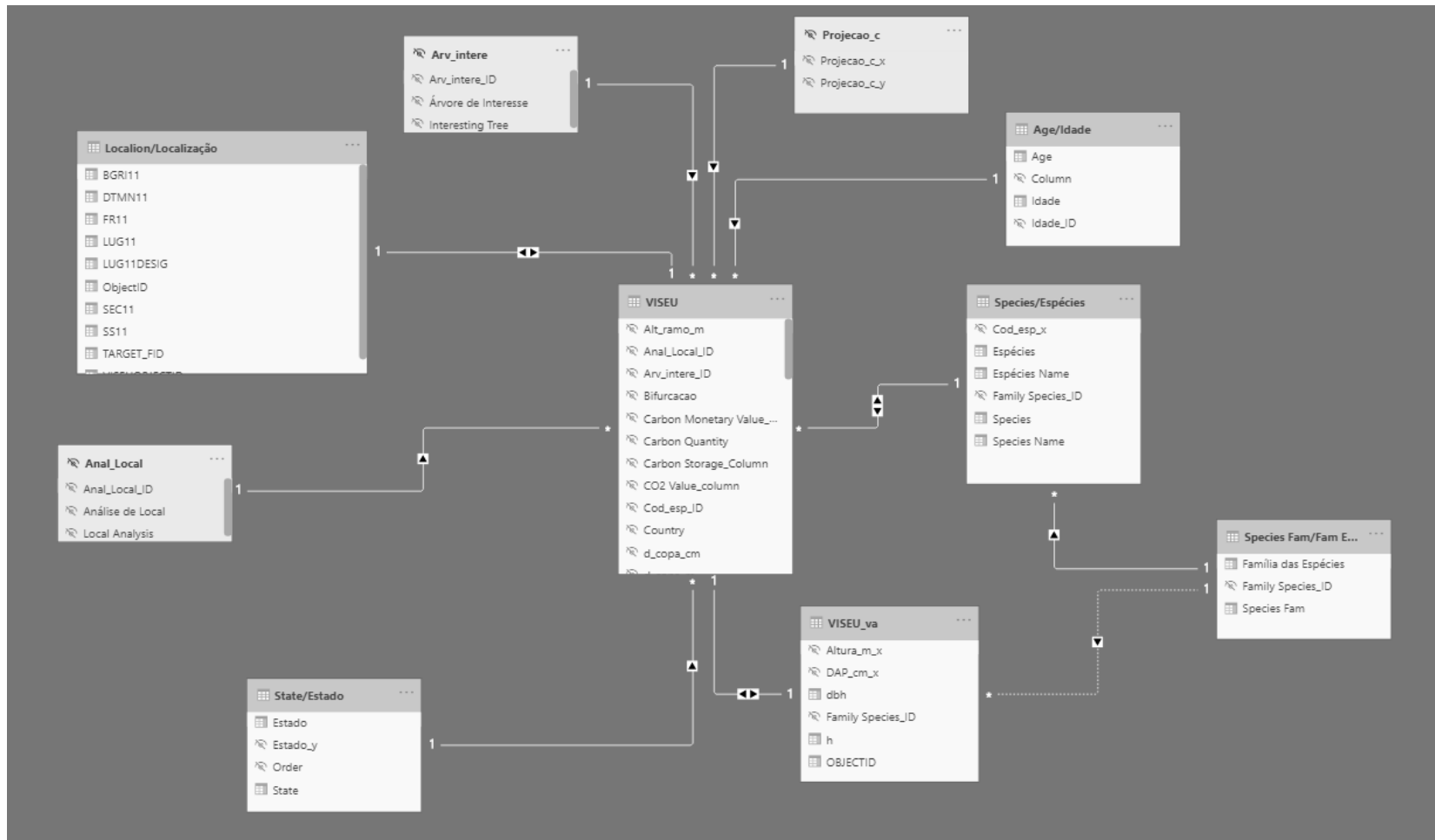


Figure 22: Data Modeler

8.5. NUMBER OF TREES AND CARBON STORAGE BY SPECIES

| Species | Nr. of trees | Carbon Stored (kg) |
|-------------------------|---------------|--------------------|
| Tilia Platyphyllos | 396 | 4.64 |
| Tilia Cordata | 435 | 5.73 |
| Betula Pendula | 226 | 33.50 |
| Fraxinus Angustifolia | 313 | 44.42 |
| Acer Saccharinum | 247 | 45.44 |
| Betula Celtiberica | 213 | 46.16 |
| Quercus Robur | 637 | 56.99 |
| Cercis Siliquastrum | 297 | 69.23 |
| Fraxinus Excelsior | 594 | 88.06 |
| Acer Pseudoplatanus | 329 | 95.53 |
| Ligustrum Japonicum | 486 | 103.25 |
| Melia Azedarach | 213 | 135.63 |
| Catalpa Bignonioides | 211 | 145.48 |
| Aesculus Hippocastanum | 275 | 230.08 |
| Liquidambar Styraciflua | 912 | 293.25 |
| Celtis Australis | 378 | 303.68 |
| Acer Negundo | 1,337 | 505.81 |
| Populus Nigra 'Italica' | 480 | 875.97 |
| Tilia Tomentosa | 1,132 | 2,572.34 |
| Others | 3,033 | 3,266.18 |
| Total | 12,144 | 8,921.39 |

8.6. HANDLING OUTLIERS

The *figure 23* and *figure 24* represent the boxplots identified that had the presence of outliers:

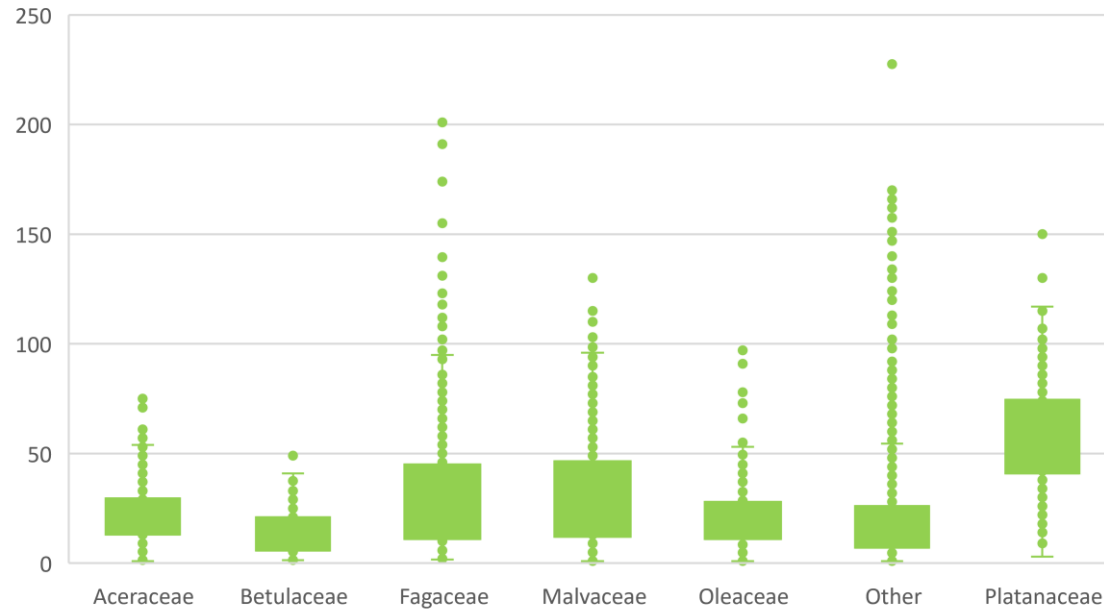


Figure 23: Dbh_{cm} distribution, by species family

The first identified variable that has the presence of outliers was the dbh_{cm} . It was decided to drop the data rows that respect the *table 10* rules.

| Family species | Rule | Deleted Data |
|-----------------------|------------------|---------------------|
| Aceraceae | $dbh_{cm} > 65$ | |
| Betulaceae | $dbh_{cm} > 45$ | |
| Fagaceae | $dbh_{cm} > 150$ | |
| Malvaceae | $dbh_{cm} > 100$ | 33 registers |
| Oleaceae | $dbh_{cm} > 85$ | |
| Other | $dbh_{cm} > 130$ | |
| Platanaceae | $dbh_{cm} > 130$ | |

Table 10: Dbh_{cm} detection of Outliers

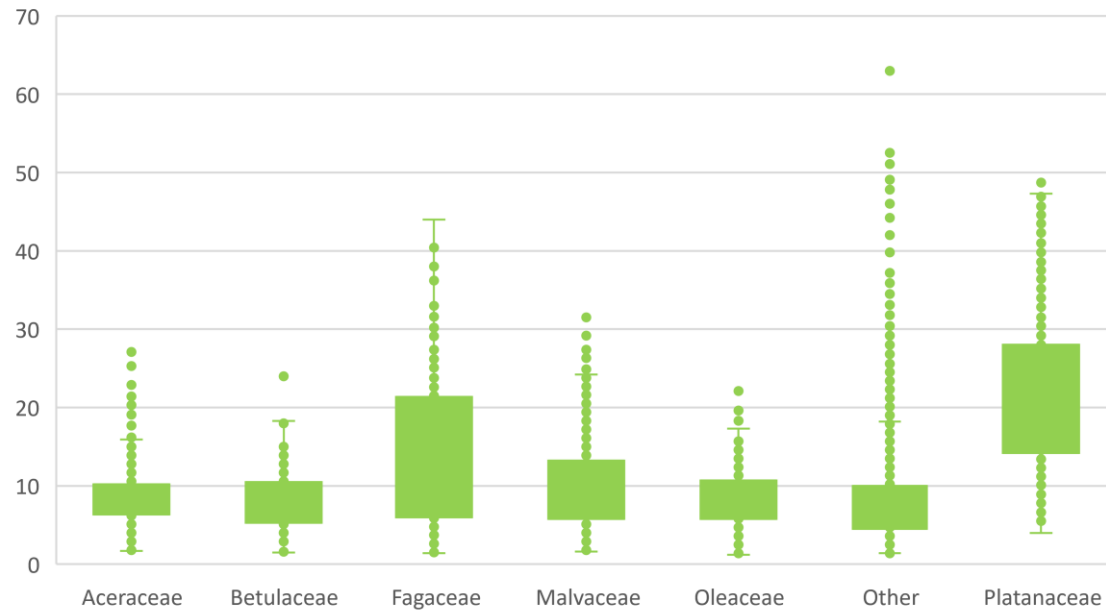


Figure 24: H_m distribution, by species family

Finally, the variable h_m was analysed. It was decided to drop the data rows that had h_m that not respect the follow *table 11* rules:

| Family species | Rule | Deleted Data |
|-----------------------|-------------|---------------------|
| Aceraceae | $h_m > 23$ | 80 registers |
| Betulaceae | $h_m > 18$ | |
| Fagaceae | $h_m > 30$ | |
| Malvaceae | $h_m > 25$ | |
| Oleaceae | $h_m > 16$ | |
| Other | $h_m > 35$ | |
| Platanaceae | $h_m > 40$ | |

Table 11: H_m detection of Outliers

8.7. DIAMETER AT BREAST HEIGHT E HEIGHT EQUATIONS IMPACT

| Allometric equation | Family species | dbh_{cm} Impact | H_m Impact |
|---|----------------|---|--|
| $V = 0.012668 + 0.0000737 \times dbh^2 \times h^{0.75}$ | Aceraceae | $\frac{\delta V}{\delta dbh} = 2dbh$ | $\frac{\delta V}{\delta h} = 0.75h$ |
| $V = 0.00004124 \times$ $(1.9302 \times \log dbh + (0.0209 \times (\log dbh)^2)) +$ $10^{(0.129 \times \log h) - (0.1903 \times (\log h)^2)}$ | Malvaceae | $\frac{\delta V}{\delta dbh} = V \left(\frac{1.9302}{dbh \ln 10} + \frac{0.0418 \log dbh}{dbh \ln 10} \right)$ | $\frac{\delta V}{\delta h} = V \left(\frac{0.129}{h \ln 10} - \frac{0.3806 \log h}{h \ln 10} \right)$ |
| $V = 0.00030648$ $\times 10^{1.2676 \times \log dbh + 0.3102 \times \log dbh^2 +$ $0.4929 \times \log h + 0.0962 \times \log h^2}$ | Oleaceae | $\frac{\delta V}{\delta dbh} = \frac{0.00133235 dbh^{0.888} h^{0.6853}}{\ln 10}$ | $\frac{\delta V}{\delta h} = \frac{0.00048361 dbh^{1.888}}{y^{0.3147} \ln 10}$ |

| Allometric equation | Family species | dbh_{cm} Impact | H_m Impact |
|---|----------------|--|---|
| $V = 8.141 \times 10^{-5} \times 10^{2.248 \cdot \log((DBH)) - 0.2062 \cdot \log((DBH))^2 + 0.1946 \cdot \log((H)) + 0.4147 \cdot \log((H))^2}$ | Betulaceae | $\frac{89.40699379 h^{1.024} dbh^{0.8356}}{100000 \ln 10}$ | $\frac{\delta V}{\delta h} = \left(\frac{0.1946}{h \ln 10} + \frac{8.294}{10h \ln 10} \right) + \frac{18.74534524 \times 10^{2.248 \log dbh - 0.2062 \log dbh^2 + 0.1946 \log h + 0.4147 \log 10h^2}}{100000}$ |
| $V = 0.00009464 \times 10^{1.9341 \cdot \log dbh - 0.0722 \cdot \log(dbh)^2 + 0.6365 \cdot \log h + 0.172 \cdot \log(h)^2}$ | Fagaceae | | |
| $V = 0.0000485 (dbh^{2.43642}) (h^{0.39168})$ | Platanaceae | $\frac{\delta V}{\delta dbh} = 0.00011816 dbh^{1.43642} h^{0.39168}$ | $\frac{\delta V}{\delta h} = \frac{0.0001899 dbh^{2.43642}}{h^{0.60832}}$ |

| Allometric equation | Family species | dbh_{cm} Impact | H_m Impact |
|---|------------------|--|---|
| $V = 0.0001967(dbh^{1.951853})(h^{0.664255})$ | General Equation | $\frac{\delta V}{\delta dbh} = 0.00038392dbh^{0.951853}h^{0.664255}$ | $\frac{\delta V}{\delta h} = \frac{0.00013065dbh^{1.951853}}{h^{0.335745}}$ |

8.8. AGE AND STATE

| Age | Nr. of trees (%) |
|----------|------------------|
| [0-10[| 31.43% |
| [10-20[| 29.85% |
| [20-40[| 20.36% |
| [40-60[| 10.19% |
| [60-80[| 5.62% |
| [80-100[| 1.80% |
| [100-∞[| 0.76% |
| Total | 100% |

Table 12: Number of trees by age

