

A Work Project, presented as part of the requirements for the Award of a Master's degree in  
Business Analytics from the Nova School of Business and Economics.

INTEGRATED DATA-DRIVEN OPTIMISATION OF SUSTAINABLE ENERGY  
DISTRIBUTION NETWORKS IN PORTUGAL

ELECTRIC MOBILITY AND GRID DEVELOPMENT

RITA BETTENCOURT DE OLIVEIRA

Work project carried out under the supervision of:

Sreyaa Guha

20/12/2023

## **Abstract**

This thesis combines several research projects focusing on sustainable energy and utility services. It examines the relationship between network connectivity and utility service continuity, the adoption of Self-Consumption Small Production Unit for Self-Consumption (UPACs) in Portugal, the expansion of Electric Vehicle (EV) Charging Points Equipment (CPEs) in municipalities, and the effectiveness of electric load forecasting models. The findings reveal a complex relationship between network requests and service continuity, a significant influence of environmental factors on UPAC adoption, complex factors driving EV infrastructure, and the superior accuracy of *SARIMAX* over *XGBoost* in forecasting. These insights provide valuable guidance for Portugal's policymaking, infrastructure development, and energy management.

**Keywords:** Energy, Service Metrics, Remote Work, Electric Vehicles, Electric Vehicle Charging Points, Zero Emissions Resilient (ZER), Population Density, Charging Points Density, Electric Mobility, Grid Connections, Renewable Energy, Small Production Unit for Self-Consumption, Grid Development, Energy Distribution, *SARIMAX XGBoost*, *Load Forecast*.

This work used infrastructure and resources funded by Fundação para a Ciência e a Tecnologia (UID/ECO/00124/2013, UID/ECO/00124/2019 and Social Sciences DataLab, Project 22209), POR Lisboa (LISBOA-01-0145-FEDER-007722 and Social Sciences DataLab, Project 22209) and POR Norte (Social Sciences DataLab, Project 22209)

## **Table of Contents**

<b>1. Introduction.....</b>	<b>1</b>
<b>2. Literature Review .....</b>	<b>3</b>
<b>3. Discussion.....</b>	<b>17</b>
<b>4. Limitations of the Study.....</b>	<b>20</b>
<b>5. Conclusion.....</b>	<b>21</b>
<b>6. Rita Oliveira - Electric Mobility and Grid Development.....</b>	<b>25</b>
<b>6.1 Introduction.....</b>	<b>25</b>
<b>6.2 Literature Review .....</b>	<b>26</b>
<b>6.3 Research Question &amp; Hypothesis .....</b>	<b>27</b>
<b>6.4 Methods.....</b>	<b>28</b>
6.4.1 Data Selection .....	28
6.4.2 Model Selection .....	29
<b>6.5 Data Analysis.....</b>	<b>31</b>
6.5.1 Data Description .....	31
6.5.2 Exploratory Data Analysis .....	32
6.5.3 Results Report.....	35
6.5.4 Predictive Models .....	39
<b>6.6 Discussion.....</b>	<b>39</b>
6.6.1 Limitations of the Study.....	41
<b>6.7 Conclusion .....</b>	<b>42</b>
<b>References .....</b>	<b>43</b>
<b>Appendix.....</b>	<b>46</b>



## List of Figures

Figure 1. Portugal National Transport Grid Map. Source: REN. ....	7
Figure 2. Electricity Generation by Source, Portugal, 2000-2021. Source: IEA.....	7
Figure 3. Overall Electricity Production Rate, Portugal, 2000-2021. Source: IEA.....	8
Figure 4. Distribution of CPEs in Portugal.....	16
Figure 5. Global Irradiation and Solar Electricity Potential vs. UPAC Installations in Portugal. Source: European Commission, E-Redes. ....	48
Figure 6. Histogram of Number of Installations.....	54
Figure 7. Installed Power Range (kW). ....	54
Figure 8. Average Installed Power by Municipality.....	55
Figure 9. . Scatter Plots Showcasing the Relationship Between the Number of Installations and Key Variables: UPAC Total Installed Power, Energy Consumption, Direct Radiation.....	55
Figure 10. Comparison of Actual vs. Predicted Values Using Linear and Lasso Regression Models.....	58
Figure 11. Average Active Energy by District. ....	89
Figure 12. Scatter Plots of Weather Variables vs. Active Energy. ....	90

## List of Tables

Table 1. OLS Regression Results – Energy Not Delivered. ....	12
Table 2. Statistical Results - Energy Not Delivered. ....	12
Table 3. OLS Regression Results - Number of Orders.....	13
Table 4. Statistical Results - Number of Orders. ....	13
Table 5. Descriptive Statistics - Number of Installations. ....	28
Table 6. Evaluation Metrics of the Models.....	31
Table 7. Lasso Regression Highest Coefficients. ....	32
Table 8. Independent Variable Importance in the XGBoost Model. ....	33
Table 9. Descriptive Statistics for Dataset 1. ....	32
Table 10. OLS Regression Results - Density of CPE per City. ....	36
Table 11. Regression Analysis Summary. ....	36
Table 12. Contingency Table. ....	37
Table 13. Classification Performance Metrics. ....	38
Table 14. Model Accuracy at Varying Threshold Levels.....	38
Table 15. SARIMA's Evaluation Metrics .....	68
Table 16. XGBoost Parameters.....	69
Table 17. XGBoost's Evaluation Metrics. ....	69
Table 18. Hybrid Model Evaluation Metrics. ....	71

## Abbreviations

<b>ACEA</b>	European Association of Automobile Manufacturers
<b>ACF</b>	Autocorrelation Function
<b>ADF</b>	Augmented Dickey-Fuller
<b>AI</b>	Artificial Intelligence
<b>AIC</b>	Akaike Information Criterion
<b>APREN</b>	Portuguese Renewable Energy Association
<b>CPE</b>	Charging Point Equipment
<b>DGEG</b>	General Direction of Energy and Geology
<b>DSM</b>	Demand-Side Management
<b>DSO</b>	Distribution System Operator
<b>EDA</b>	Exploratory Data Analysis
<b>EESC</b>	European Economic and Social Committee
<b>END</b>	Energy Not Delivered
<b>EU</b>	European Union
<b>EV</b>	Electric Vehicle
<b>FEV</b>	Full Electric Vehicles
<b>Gw</b>	Gigawatt
<b>GWh</b>	Gigawatt Hour
<b>IEA</b>	International Energy Agency
<b>kV</b>	Kilovolt
<b>kW</b>	Kilowatt
<b>MAE</b>	Mean Absolute Error
<b>MAIFI</b>	Momentary Average Interruption Frequency Index
<b>MAPE</b>	Mean Absolute Percentage Error
<b>ML</b>	Machine Learning

## Group Part

<b>MSE</b>	Mean Squared Error
<b>MW</b>	Megawatts
<b>OLS</b>	Ordinary Least Squares Regression
<b>PACF</b>	Partial Autocorrelation Function
<b>PEV</b>	Plug-in Electric Vehicle
<b>PV</b>	Photovoltaic
<b>REN</b>	National Energy Network
<b>RESP</b>	Public Service Electric Network
<b>RMSE</b>	Root Mean Squared Error
<b>RRP</b>	Recovery and Resilience Plan
<b>SAIDI</b>	System Average Interruption Duration Index
<b>SAIFI</b>	System Average Interruption Frequency Index
<b>SARIMAX</b>	Seasonal Autoregressive Integrated Moving Average with Exogenous Factors
<b>TIEPI</b>	Total Energy Index of Power Interruptions
<b>UPAC</b>	Small Production Unit for Self-Consumption
<b>XGBoost</b>	Extreme Gradient Boosting
<b>ZER</b>	Zero Emissions Resilient

## **1. Introduction**

This work project explores the complex aspects of utility services and sustainable energy in a world where technological advancement and sustainability concerns drive change. This collaborative effort brings together various research projects that explore different but related facets of energy systems and their significant influence on society.

The first stage in this investigation thoroughly examines how network connection requests affect utility service continuity. In an age where utility networks are growing increasingly complex and networked, it is critical to comprehend the relationship between these requests and service provision. This study explores how changing utility service management affects the energy industry and highlights its difficulties, such as increased remote work. The relationship between network connection requests and service continuity metrics offers insights into the demands on utility infrastructures, highlighting the critical need for resilient and adaptive systems.

In tandem with this, this work project investigates the critical role of renewable energy, concentrating on Portugal's UPACs. Adopting UPACs is a significant step toward improving sustainability and decentralizing energy production. This section of the study investigates the factors affecting how UPACs are implemented in Portuguese municipalities and provides a thorough understanding of the regional differences that influence the adoption of renewable energy by focusing on environmental, meteorological, and consumption influences. Prominent discoveries highlight the significance of meteorological factors, such as exposure to sunlight, which significantly influence the rate of UPAC installations.

The evolution of transportation through integrating Electric Vehicles (EVs) and related infrastructure is scrutinized. The research delves into the rise of Charging Points Equipment (CPEs) throughout Portuguese municipalities, marking an essential facet of mobility's progression. Moreover, the investigation delves into economic, demographic, and environmental influences on EV infrastructure growth. The analysis sheds light on the decisive

## Group Part

role urban planning and policy play in sculpting CPE distribution and density, unraveling the complex interplay of these elements.

Continuing the exploration, the focus shifts to electric load forecasting, which is essential to contemporary power systems' functioning. Reliable load forecasting ensures that energy distribution networks can effectively supply demand. The study advances this field by evaluating the predictive power of sophisticated statistical and machine learning (ML) models applied to the Portuguese energy sector. We examine the efficacy of the eXtreme Gradient Boosting (*XGBoost*) and Seasonal Autoregressive Integrated Moving Average with Exogenous Factors (*SARIMAX*) models and a hybrid approach that combines these techniques. This section of the work project incorporates meteorological factors, like temperature and precipitation, to improve the forecasting precision of energy consumption.

Various methodological techniques and reliable data sources support the group's investigations. Each segment of the thesis is informed by a tailored mix of analytical methods, including ML algorithms, *Granger* causality tests, and regression analyses, to thoroughly address the unique research questions and hypotheses within each section. The E-Redes Open Data Portal provides a wealth of data on utility services and energy metrics, establishing a solid empirical basis for the research. This deliberate and methodical data gathering, and analysis not only supports the study's relevance to real-world applications but also strengthens the results' trustworthiness.

The research uncovers complex dynamics within the utility sector, revealing a correlation between network connection requests and remote work order volumes, with minimal effects on service continuity indicators. It showcases the challenges service systems face with changing demands. Regarding renewable energy, the study emphasizes the pronounced impact of weather on the uptake of UPACs in Portugal, advocating for localized policy and planning. EV CPE distribution in electric mobility varies significantly across municipalities, shaped by localized economic, demographic, and environmental factors.

## Group Part

Finally, exploring electric load forecasting highlighted the superior accuracy of *SARIMAX* models compared to *XGBoost*. Introducing a hybrid model showed additional improvements, signifying the advantages of combining different methodologies in predicting energy system demands.

The work project offers a thorough analysis of utility service management and sustainable energy systems, shedding light on the sector's prospects and obstacles by weaving together studies on load forecasting, renewable energy uptake, electric mobility infrastructure, and network connectivity. It aspires to enrich the field, providing insights for academia, industry, and policy formulation. The intent is to forge a knowledge base to inform policy crafting, infrastructure expansion, and service refinement, steering toward greater resilience, sustainability, and efficiency in the utility industry.

## **2. Literature Review**

The costs associated with energy consumption, both financially and environmentally, are currently a major issue in our society. Renewable energy consumption was not previously a problem, but as society has grown more environmentally conscious, this idea is beginning to shift. The use and exploitation of renewable energies is now necessary due to the rise in energy demand brought on by the world's population growth and the fast expansion of emerging economies, which is also connected to the depletion of non-renewable resources. Considering the possibility that it will help avert an impending energy crisis and preserve Earth's ecology, we must keep funding its advancement.

Portugal has committed itself to the ambitious objective of attaining carbon neutrality by the year 2050, aligning with global and European targets established in implementing the Paris Agreement. This commitment entails a substantial reduction of over 85% in greenhouse gas emissions, measured against the benchmark of 2005, coupled with the imperative of fostering a carbon sequestration capacity of approximately 13 million tons.

## Group Part

As the General Direction of Energy and Geology (DGEG) articulated, the anticipated energy transition in the coming decade necessitates an investment exceeding 25000 million euros. This undertaking requires a sophisticated coordination of collective wills, aligning policies, incentives, and financing mechanisms. To facilitate this transition effectively, the mobilization of a comprehensive suite of legal and planning instruments becomes imperative. These instruments serve as the strategic framework for achieving a tangible reduction in emissions while concurrently fostering investment, employment, and innovation. The DGEG emphasizes the crucial role of these measures in steering Portugal through the complexities of the energy transition and ensuring its success.

In pursuing comprehensive decarbonization, DGEG underscores its multifaceted significance as a strategic driver for investment and employment creation. The commitment to spearheading the energy transition is resolute, particularly through substantial investments in renewable production. As articulated by the DGEG, there is an imperative to more than double the installed capacity of renewable energy within the coming decade, with the ambitious goal of surpassing 80% of renewables in electricity production.

The outlined targets for Portugal by 2030 further emphasize the nation's dedication to sustainable practices. Aiming for 47% renewable energy in gross final energy consumption and a parallel target of 20% renewable energy in the transport sector by the specified timeline demonstrates a proactive stance toward achieving a cleaner and more sustainable energy landscape.

According to DGEG's perspective, the coming decade emerges as a critical juncture demanding heightened efforts to substantially curtail greenhouse gas emissions. This commitment entails the adoption of ambitious decarbonization targets, coupled with a deliberate focus on the integration of renewable energy sources and the enhancement of energy efficiency. The DGEG underscores the importance of ensuring a fair and inclusive transition,

## Group Part

recognizing the social and economic dimensions as indispensable conditions for the success of this transformative vision.

To accelerate economic decarbonization, the XXIII Constitutional Government's program pledges various commitments, including:

- Developing a five-year carbon budget that specifies a multi-year horizon, expediting the execution of the 2030 National Energy and Climate Plans and the 2050 Carbon Neutrality Roadmap, supporting regional carbon neutrality roadmaps, providing methods for evaluating the impact of legislation on climate action, and eliminate administrative barriers that result in excessive context costs without environmental added benefit.
- Putting into practice the energy efficiency investment anticipated under the Recovery and Resilience Plan (RRP); 300 million euros for energy efficiency in residential structures, with a focus on lower-income families; 310 million euros for energy efficiency in the business sector and Public Administration service buildings; 715 million euros in the decarbonization of industry and 185 million euros in investment in hydrogen and renewable gases.
- Raise the capacity of solar energy production by at least 2 gw, keep holding auctions for new power plants, and encourage and assist the development of energy communities and self-consumption.
- Encourage green bond issuance, support micro-credit platforms for low-carbon investments, and facilitate collaboration between the Fund for Innovation, Technology, and Circular Economy and the Environmental Fund to back decarbonization projects and enhance resource efficiency.

According to the IEA (International Energy Agency), in 2020, Portugal's electricity system was divided into two parts: a distribution system with high, medium, and low voltage lines and cables run by 13 distribution system operators (DSOs) and a transmission system with very-

## Group Part

high voltage lines connected to Spain through nine cross-border interconnections. The transmission network has recently been enhanced to integrate new renewable generation. The national transmission network measured 9 036 km by the end of 2020, with 2 711 km operating at 400 kV, 3 780 km operating at 220 kV, and 2 545 km operating at 150 kV. The installed transformer capacity of the network was 38 463 megavolt-amperes.

Eleven of Portugal's thirteen DSO in 2020 were located on the country's mainland. 99.5% of customers in Portugal are connected at low voltage, and E-Redes is the only DSO for high-voltage and medium-voltage distribution systems. It also manages low-voltage distribution systems in 278 of the 308 municipalities in Portugal. The remaining 0.5% of consumers are served by ten more small-scale DSOs that run low-voltage distribution networks at the municipal level. Furthermore, the DSOs of the islands of Madeira and Azores are separate from the rest of the value chain.

To carry more renewable electricity to consumption centers, the transmission network has been strengthened in recent years due to the integration of high levels of new renewable generation. Additionally, the capacity of interconnection between Portugal and Spain has increased due to agreements made between the two nations to enhance the Iberian power market. Portuguese import capacity climbed from 1,112 MW to 2,970 MW in 2020, while export interconnection capacity went from 1,183 MW in 2010 to 2,925 MW in 2020. In Figure 1, it is possible to observe the National Transport Grid Map for the year 2020, as presented in the National Energy Network (REN).

## Group Part

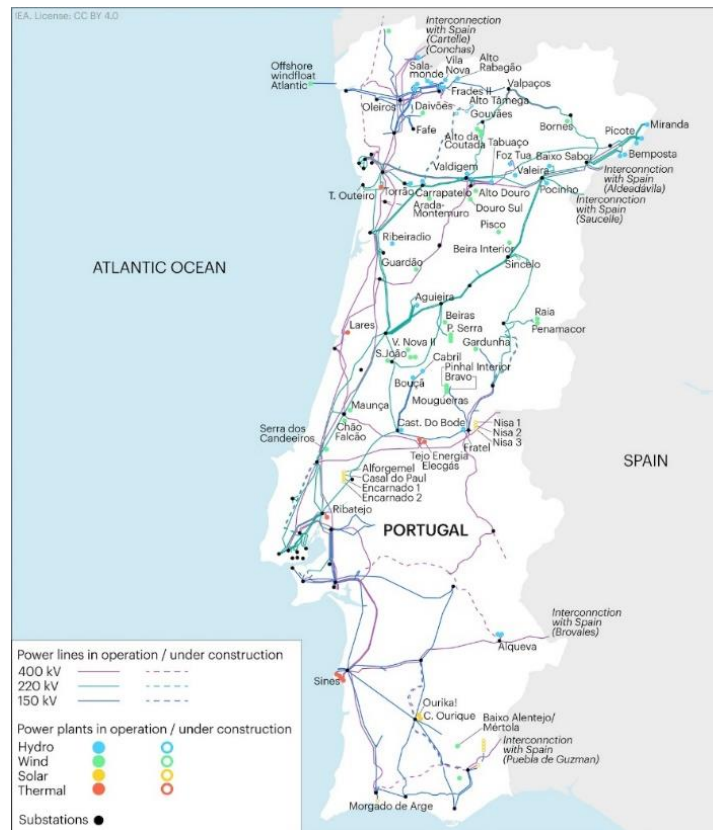


Figure 1. Portugal National Transport Grid Map. Source: REN.

According to the IEA, Portugal's electricity provision exhibits a division between renewable sources, predominantly wind and hydro, and fossil fuels, primarily natural gas and coal (Figure 2).

Electricity generation by source, Portugal, 2000-2021

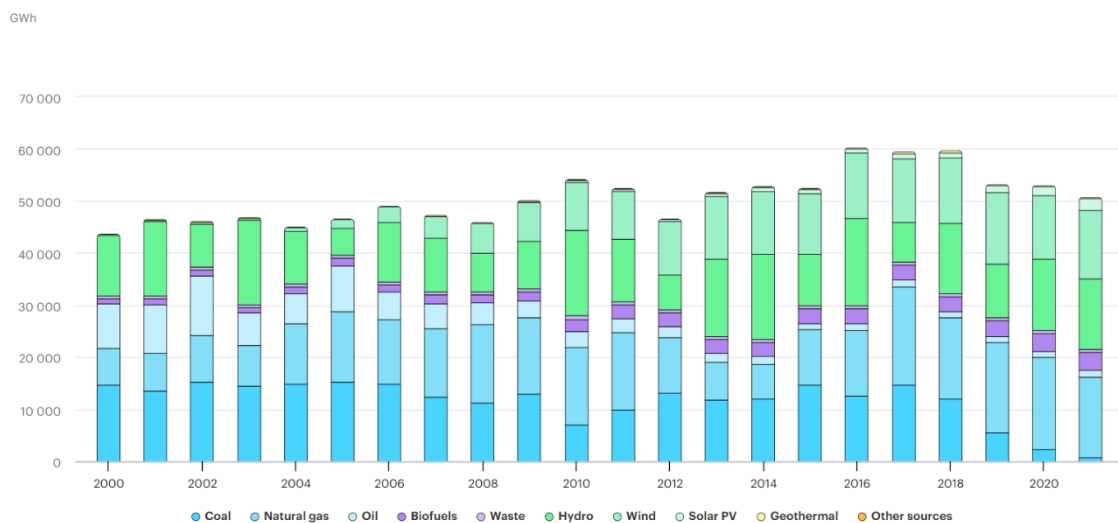


Figure 2. Electricity Generation by Source, Portugal, 2000-2021. Source: IEA.

## Group Part

In 2016 Portugal achieved a milestone by transitioning into a net electricity exporter, a trend sustained until 2019. Notably, this shift was attributed to the expansion of renewable energy generation. However, in the subsequent year, Portugal reverted to being a net importer of electricity (Figure 3).

Overall electricity production rate, Portugal, 2000-2021

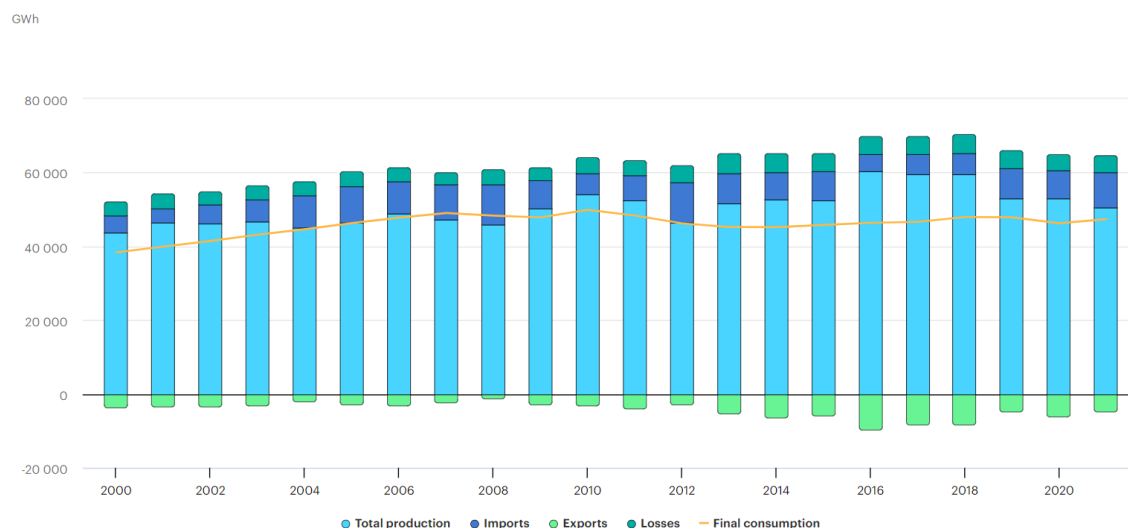


Figure 3. Overall Electricity Production Rate, Portugal, 2000-2021. Source: IEA.

In 2023, Portugal achieved a groundbreaking milestone in renewable energy production, setting a new record lasting almost a week. According to data from REN, from 4 AM on October 31<sup>st</sup> to 9 AM on November 6<sup>th</sup>, spanning 149 consecutive hours, Portugal produced 1102 gw per hour of renewable energy. This continuous output was more than sufficient to meet the combined industrial and residential consumption demands during the same period, totaling 840 GWh.

This accomplishment surpasses the previous record set in 2019, which stood at 131 hours of continuous renewable energy production. REN also highlights two additional milestones during this record-setting period. Between 10 AM on November 1<sup>st</sup> and 9 AM on November 5<sup>th</sup>, spanning 95 consecutive hours, Portugal's renewable production exceeded consumption without natural gas plants. During this time, Portugal even exported excess energy to Spain, surpassing the previous export record of 52 hours set in 2018.

## Group Part

Furthermore, from 10 PM on October 31<sup>st</sup> to 9 AM on November 6<sup>th</sup>, renewable energy production exceeded the national electrical system's needs, including the pumping requirements in hydroelectric reservoirs. Notably, hydroelectric production achieved these remarkable volumes significantly, showcasing Portugal's strides in sustainable and resilient energy systems.

Pragash, K, and P (2023) says that improved energy efficiency, less carbon emissions, better resilience and dependability, and enhanced system flexibility are some of the main attributes and advantages of smart grid technology. One major problem is integrating these variable energy sources into the current electrical infrastructure. The conventional electrical grid was intended to manage one-way power flows from the power plant to the end users. However, because renewable energy sources are spread geographically and produce power sporadically, they may cause grid oscillations and raise the possibility of blackouts. The following are a few of the major obstacles to incorporating renewable energy into the electrical grid:

- Intermittency and variability: The nature of renewable energy sources, like wind and solar power, is intermittent and variable, meaning that the amount of energy they produce varies according to the weather and other circumstances.
- Grid balancing and stability: The addition of significant amounts of intermittent renewable energy may impact these aspects of the system. To keep the system stable, grid managers must be able to swiftly balance supply and demand when the production of renewable energy sources varies.
- Economic challenges: Including the need for new price structures and incentives to promote the development of renewable energy projects, integrating renewable energy sources into the grid may provide economic challenges. Legislators and regulators must get involved to establish a supportive regulatory framework for the growth and integration of renewable energy.

## Group Part

- Enhancing adaptability and responsiveness is essential to incorporate renewable energy sources into the grid successfully. The integration of renewables significantly influences grid operations and management, with critical impacts including:
- Lower carbon emissions: Using renewable energy sources like wind and solar can help lower carbon emissions and slow global warming. One of the main advantages of incorporating renewable energy into the electrical system is this.
- Intermittency and variability: The nature of renewable energy sources is frequently intermittent and variable, which means that the weather and other variables can affect how much energy they produce.

The transformative impact of data-driven techniques in energy distribution is paving the way for smarter, more efficient electricity networks. At the heart of this transformation are smart grids, which employ big data analytics to enable real-time monitoring and dynamic management of energy flows. This optimizes the distribution system's performance by utilizing data to forecast demand, integrate renewable sources effectively, and manage operations adaptively.

Ahmad et al. (2022) underscores the significance of advancements in ML technologies for optimizing decision-making within energy networks. These technologies are at the forefront of the shift towards a more integrated approach to energy distribution, driven by the urgency to address climate change and adhere to sustainability policies. The influence of ML is widespread, forging new paths in energy distribution across numerous areas, from the development of advanced energy materials to the strategic planning of energy systems, enhancing storage device capabilities, and bolstering energy efficiency.

Moreover, ML's reach extends beyond production. It plays a pivotal role in predictive maintenance, showcasing the extensive scope of ML in revolutionizing the sector. Data-driven methodologies are also redefining the management of energy outages. Big data analytics are indispensable for renewable energy and microgrid management, where precise and efficient

## Group Part

power generation forecasting - rooted in extensive weather data - is critical. Such analytical prowess is essential for augmenting the efficacy of asset management and cooperative operations within the power industry, thereby enhancing grid reliability and stability.

Within the demand-side management (DSM) domain, big data analytics offers many insights extracted from large-scale electricity usage data, supporting the development of demand-side solutions and marketing plans. With historical load patterns, meteorological data, and societal trends contributing to a more nuanced understanding of future demands, load forecasting emerges as a critical research domain within intelligent grids.

Drawing insights from Crispim et al. (2014) in their paper on "Smart Grids in the EU with Smart Regulation: Experiences from the UK, Italy, and Portugal," the implementation of smart grids across the European Union (EU) offers a wealth of informative case studies. According to Crispim et al. (2014), the EU's adoption of smart grid solutions represents a collaborative effort towards cooperative development, as embodied in the EU 7th Framework Programme and Horizon 2020. These initiatives underscore the EU's commitment to integrating renewable energy sources, advancing demand response strategies, and optimizing the usage of electricity in new technologies like EVs and heat pumps.

Furthermore, the transition to smart grids in the EU, as expounded by Crispim et al. (2014), is influenced by several key factors. These include the urgency to lower greenhouse gas emissions, the challenges associated with managing grids that handle variable and unpredictable generations, and the integration of new categories of energy consumers, such as EVs. This shift is also pivotal in redefining regulatory practices and shaping the market dynamics in various EU member states.

Building upon the insights from Crispim et al. (2014) regarding the implementation of smart grids across the EU, Portugal presents a case study of significant advancements in smart grid development. Soares et al. (2015) provide a detailed account of Portugal's major smart grid projects resulting from collaboration between academia and industry. These projects, namely

## Group Part

InovGrid and REIVE, showcase the implementation of advanced metering systems, the integration of distributed energy resources such as EVs, microgeneration units, and energy storage devices into low-voltage networks, and the development of new functionalities for system operators based on microgrid concepts.

The InovGrid project aimed to develop an advanced metering infrastructure and functionalities that would enhance the intelligence of the distribution grid. This included managing and controlling microgeneration in low-voltage grids and providing conditions for customers to access new services. The project's real-world application was demonstrated in Évora, designated as InovCity, where the developed devices and control architecture were tested.

The REIVE project focused on the technical and market integration of renewables and EVs, emphasizing the role of EVs as flexible resources that can provide load flexibility and storage capacity. The project envisioned new grid-supporting functionalities and developed a market structure where an aggregator entity could provide market representation for the integrated EVs and microgeneration units.

Soares et al. (2015) also highlight the development of a Smart Grids and Electric Vehicle Laboratory as an outcome of the REIVE project. This laboratory is a testbed for evaluating smart grid concepts, control algorithms, communication solutions, and information exchange schemes. It has been instrumental in pre-validating functionalities intended for implementation in InovCity, including voltage and frequency regulation strategies and coordinated operation of microgeneration, storage, and EV charging.

This study also underlines the practical applications and benefits of smart grids. The InovGrid and REIVE projects not only facilitated the integration of renewable energy sources and EVs into the energy network but also demonstrated how smart grid technologies can enhance energy efficiency, promote consumer engagement in energy management, and contribute to the overall sustainability goals of the EU.

## Group Part

Remote work orders are usually issued in the energy distribution industry when system faults or disruptions need to be fixed. The consistency and dependability of the service provided can have a significant impact on how frequently these work orders are handled. By combining data-driven technologies and smart grids, service continuity can be improved, and the number of remote work orders can be decreased.

Big data analytics and ML enable smart grids to anticipate faults or disruptions and identify those areas in advance. With proactive maintenance and repairs made possible by this predictive capability, the number of remote work orders may be decreased. Using ML algorithms, historical data can be analyzed to spot trends or warning indications of impending failures, enabling prompt intervention to stop problems before they become more serious.

Furthermore, data-driven methods can enhance how remote work orders are handled when they arise. Through the analysis of data from multiple sources, including historical maintenance records, weather patterns, and real-time grid performance, utilities can more effectively allocate resources and prioritize work orders based on urgency. This can reduce the impact on consumers by resulting in quicker response times and more efficient issue resolution.

The European Economic and Social Committee (EESC) emphasizes the crucial symbiosis of energy self-production and self-consumption in the context of environmental protection, climate change mitigation, and addressing energy poverty. Unlike fossil fuels or nuclear energy, renewable technologies inherently possess a local dimension intimately tied to their geographical locations. Predominantly comprising photovoltaic (PV) solar and wind power, supplemented by small-scale hydropower, these technologies contribute to the revitalization of self-consumption in electricity, involving the direct utilization of locally generated energy.

The evolving landscape includes potentially incorporating a local variant of future green hydrogen. Over recent years, European and national legislation, particularly in select countries, has championed self-consumption. This support encompasses individual consumption through technologies like PV panels on rooftops and collective consumption facilitated by energy

## Group Part

communities, local authorities, cooperatives, and similar entities deploying PV or wind farms. The clean and increasingly cost-effective energy produced, distributed, and consumed through these means holds significant promise for impacting accessibility and prices. Positioned as a valuable tool, it addresses energy poverty and contributes to the global effort against climate change.

Portugal reached a historic milestone in solar energy production in May 2022, becoming the fourth European nation to dramatically increase the percentage of renewable energy sources used in power production in the first five months of the year. Throughout the first half of 2022, the residential sector saw significant growth, suggesting a growing trend among families becoming more aware of the financial and environmental benefits of solar solutions. The use of government incentives contributes to this increase.

According to Mordor Intelligence, the need for an alternative electricity source and the desire to reduce the risk of climate change, in addition to the anticipated savings that are now more effective due to technical advancements, are the main reasons for this adoption by individuals.

The development of various technologies, such as EVs, is being encouraged by the decarbonization strategies devised by the EU and other European countries (Wee, Coffman, and La Croix 2018). Combined with low-carbon energy generation, they offer a viable way to lessen the transportation sector's substantial reliance on oil and gas products. Europe, along with China and the US, is one of the biggest EV markets in the world, with sales significantly growing year over year. However, due to various obstacles, such as increased investment costs, range anxiety, and a lack of adequate charging infrastructure, EVs still only make up a small portion of the global auto market in most countries.

The quantity of plug-in electric vehicles (PEVs) is impacted by government policy about PEVs. Three primary elements comprise government policies: incentives, regulations, and technology-related policies. Tax breaks and financial aid are two ways to encourage the

## Group Part

purchase of PEVs. Numerous studies have shown a direct correlation between PEV sales and incentives. According to Sierzchula et al. (2014) the federal tax credit financed 30% of PEV sales.

The Portuguese government has been encouraging the development of electric mobility solutions and the use of Full Electric Vehicles (FEV). With the goal of reducing CO<sub>2</sub> emissions by 25% globally by 2020 in Portugal the Portuguese Electric Mobility Program (Mobi.E) had a partnership of enterprises and a network of municipalities to install and operate a pilot to promote electric mobility throughout the country. This initiative defined a market model that was piloted and developed in two later phases.

Portugal ranks among the top five European countries regarding charging points per 100 kilometers of road. According to ACEA (European Association of Automobile Manufacturers) research, Portugal ranks fourth on this statistic, trailing only the Netherlands, Luxembourg, and Germany. Furthermore, in May 2023, the share of EV sales was around 16%, compared to around 4.5% in Spain or Italy.

These are significant milestones, but they will need to be expanded to cover the entire region and the predicted growth in demand as the entire automotive industry is compelled to undergo an unprecedented electrical conversion to meet climate goals. Furthermore, the European Commission has set a maximum distance of 70 kilometers between two roadside charging locations.

The Portuguese problem stems mainly from more than 60% of homes needing a garage and relying only on the public network. On the other hand, the regional disparity in the distribution of this service has been identified as a weakness (Figure 4). According to *Mobi.E* data, in 2023, there will be 53 sockets for every 100 kilometers of highway in Portugal and 72 for every 100,000 residents.



## Group Part

The shift towards incorporating ML, exemplified by models like *XGBoost*, represents a leap in managing large-scale, non-linear data crucial for responding to the dynamic nature of contemporary energy demands. These models excel in processing vast quantities of data and uncovering complex relationships, a necessity for modern energy distribution systems that adapt to real-time changes and diverse energy inputs.

Integrating advanced forecasting models like *SARIMAX* and *XGBoost* into energy distribution systems is crucial for enhancing optimization strategies. Such integration leads to smarter grid management, allowing dynamic adjustments in energy supply based on accurate forecasts, improving efficiency, reducing waste, and better incorporating unpredictable renewable sources. Exploring these models, especially in hybrid forms, is critical to advancing energy distribution optimization, underpinning intelligent systems that ensure stable, sustainable supply amidst increasing demand and environmental sustainability goals. These techniques are fundamental in the data-driven optimization of energy systems, providing the essential intelligence for resilient and adaptable power infrastructures.

### **3. Discussion**

This comprehensive analysis explores the intricate dynamics between executed network connection requests and their impact on service continuity indicators, remote work orders, and electric mobility inclusion. The findings indicate that increased network connection requests do not uniformly translate to improved service continuity indicators. Despite improvements in some indicators, others, such as the System Average Interruption Duration Index (SAIDI) and Energy not Delivered (END), show deterioration, highlighting a complex interplay of components. This complexity challenges existing models that predict a direct relationship between infrastructure investment and improved service outcomes.

The correlation between *network connection requests* and *end mt mwh* suggests that increased requests strain existing infrastructure or pose management challenges, affecting

## Group Part

energy delivery. This observation has profound implications for utility companies, emphasizing the need to assess the quantity and quality of infrastructure investments.

Remote work orders data corroborates the importance of strategic planning in resource allocation and labor management. The increase in remote work orders with more network connection requests demonstrates a direct relationship between network expansion and the demand for associated services.

Simultaneously, the study delves into environmental and meteorological factors, particularly their influence on UPAC adoption rates. The positive coefficients for meteorological variables, highlighted by their feature importance in models like *XGBoost*, underscore the significance of environmental conditions like sunlight exposure. This finding is pivotal for policymakers, indicating that meteorological factors are crucial for accurately predicting UPAC adoption.

Comparing the Mean Squared Error (MSE) values across different supervised ML models in hypothesis testing reveals insights into their effectiveness in predicting UPAC adoption. The varying levels of predictive accuracy, especially the significant difference indicated by the *Ridge Regression* model, suggest complexities in the dataset that some models capture better than others. These insights are invaluable for developing and implementing policies promoting UPAC adoption, with models like *Lasso Regression* offering robust frameworks for formulating targeted interventions.

The statistical analysis further examines the correlation between population density and the density of CPEs in Portuguese municipalities. Supported by both *Pearson* and *Spearman* Correlation analyses, a strong positive relationship is evident. However, the K-Means algorithm's clustering analysis reveals a nuanced picture: different clusters show varied relationships between population density, CPE density, and EV infrastructure development.

This analysis also challenges the assumption that higher income levels correlate with greater CPE density. *Regression analysis* demonstrates that population density is a significant

## Group Part

predictor, but average income per city is not, contradicting previous hypotheses. Diagnostic tests reveal a more complex relationship than a linear model can capture, suggesting the need for a broader array of factors in examining EV infrastructure placement.

Additionally, the study examines the impact of Zero Emissions Resilient (ZER) urban policies. The correlation heatmap indicates a weaker correlation between ZER policy zones and CPE density, but further categorization into restricted and non-restricted zones reveals that restricted zones have higher CPE density. This supports the hypothesis that regulatory measures might encourage CPE installation.

*Predictive modeling* indicates that city size and population density account for about 50% of the variance in CPE density, hinting at unexplored variables that might influence this phenomenon. This moderate endorsement of the hypothesis suggests a complex interplay between urban policies, population density, and CPE infrastructure development.

The analysis also addresses the relationship between voltage level and electric mobility inclusion. *Cross-validation* confirms the significant association between these factors, indicating a genuine pattern rather than a sample-specific observation. *Logistic Regression* analysis further supports this, with models showing strong predictive capabilities and indicating a higher likelihood of electric mobility inclusion at low-voltage unique installations.

Model tuning with *GridSearchCV* optimizes the logistic regression model's hyperparameters, favoring a simpler model to mitigate overfitting risk. Sensitivity analysis confirms the model's robustness, suggesting that the findings are not influenced by the data proportion used for training or testing.

Lastly, the comparative analysis of forecasting models about *SARIMAX* and *XGBoost* models reveals that *SARIMAX* models demonstrate superior predictive capabilities. Lower MSE, Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE) values and higher R-Squared and custom accuracy rates support this. A hybrid approach combining *SARIMAX*

and *XGBoost* offers a marginal improvement in forecast accuracy, confirming the effectiveness of a hybrid modeling approach in this context.

This study comprehensively explains the factors influencing service continuity, UPAC adoption rates, CPE infrastructure, and electric mobility inclusion. It highlights the complexity of these relationships and the importance of considering various factors for effective policy and infrastructure development in sustainable energy and technology adoption.

#### **4. Limitations of the Study**

In the limitations section of this thesis, several key constraints encountered during the research process need to be acknowledged, spanning various aspects of the study.

Primarily, the limitations are tied to the data and its scope. While the E-Redes statistics used in the study are extensive, they are confined to Portugal's electrical grid. This geographical and operational specificity implies that the findings might only be readily generalizable to other countries or energy sectors with further examination. Additionally, the reliance on secondary data introduces potential restrictions related to how the data was initially collected, including any inherent biases or inaccuracies in the data recording process. Given the rapidly evolving nature of the utility industry, driven by technological advancements and regulatory changes, the historical data may only partially capture the current dynamics of the sector.

Another limitation arises from the computational constraints that led to the use of district-level data despite having access to more detailed municipality-level data from E-Redes. This decision potentially limits the detail and insights that the forecasting models could have offered. Furthermore, if the data were a more extensive set of historical records, conducting a detailed municipality-by-municipality analysis would have enabled an understanding of specific local factors driving solar panel adoption. This approach could have provided more tailored and practical guidance to each municipality.

Moreover, the study faced challenges in exploring certain variables due to the need for comprehensive information and temporal data. These data availability issues hampered efforts

## Group Part

to incorporate alternative data types like population density, CPEs locations, and income levels. The absence of longitudinal data mainly restricted the ability to perform a thorough temporal analysis, which would have been crucial for more profoundly understanding trends and patterns over time.

While the *Random Forest* model demonstrated superior predictive capability, it is a complex "black box" model that presents challenges in fully understanding its workings. Although significant, the insights derived regarding the importance of features should be approached with caution as they do not fully unravel the causal mechanisms at play.

In conclusion, the scope of this study's findings and conclusions is inherently influenced by these data limitations, affecting the generalizability and depth of the analysis. Future research in this domain would benefit significantly from access to broader datasets, including time-series data, to address these limitations and further build upon the foundational research presented in this thesis.

## 5. Conclusion

In concluding this research project, it is pivotal to revisit its core thesis: "*Integrated Data-Driven Optimization for Sustainable Energy Distribution Networks in Portugal*". This central theme has guided an in-depth exploration into how data-driven strategies can revolutionize Portugal's energy distribution networks, emphasizing sustainability and efficiency. Through comprehensive analyses and discussions, the research has illuminated the critical role of technological advancements and innovative methodologies in shaping the future of energy distribution in a way that aligns with environmental sustainability and operational efficiency.

The project delved into several critical areas of Portugal's energy sector. It explored utility services, emphasizing the relationship between network demand and service continuity, revealing challenges in managing growing demands. In renewable energy, particularly UPACs, the study highlighted the crucial influence of environmental factors on adoption rates. The research on electric mobility infrastructure revealed the diverse factors influencing the

## Group Part

distribution of EV Charging Points. Furthermore, in load forecasting, the project demonstrated the efficacy of *SARIMAX* models over others. These findings collectively underline the complexities in sustainable energy management, emphasizing the need for nuanced, data-driven policy-making and strategic infrastructure development in Portugal.

In the area of Service Continuity and Remote Services, the research highlighted the intricate relationship between infrastructure growth and service continuity, presenting complex challenges in utility service management. It underscored the need for utility companies to invest in expanding and improving their networks' resilience and efficiency. Additionally, the positive correlation between network connection requests and the increase in remote work orders reflected a responsive service sector adapting to changing operational paradigms and evolving customer needs. This part of the study calls for a reevaluation of current practices, suggesting a shift towards more adaptable and efficient utility service frameworks.

The research on Renewable Energy Integration Network revealed the critical role of environmental and meteorological factors in the adoption of UPACs in Portugal. The study's findings indicate the necessity for regionally specific renewable energy policies that consider local environmental conditions. This area of research offers a comprehensive understanding of the dynamics affecting UPAC adoption, providing valuable insights for policymakers and energy managers to propel sustainable energy strategies that align with Portugal's unique environmental landscape.

In the Electric Mobility and Grid Development section, the research provided robust evidence supporting the correlation between urban density and the distribution of EV CPEs in Portuguese municipalities. This study goes beyond simple correlations, shedding light on how economic, policy, and technical factors interplay in infrastructure deployment. It challenges traditional perspectives on income and policy incentives, underlining the complexity in developing EV infrastructure. This segment emphasizes the need for multifaceted urban

## Group Part

planning and policy formulation to effectively deploy EV charging infrastructure within the sustainable transport landscape.

Finally in the Accuracy of Electric Load Forecasting Models section, the research assessed *SARIMAX* and *XGBoost* models' performance in predicting electric load, emphasizing the importance of weather variables and seasonal patterns. The *SARIMAX* model outperformed *XGBoost*, but a hybrid approach combining both models showed slight improvements in accuracy. These findings suggest the potential of integrated modeling methods in enhancing energy forecasts, crucial for optimizing sustainable energy distribution in Portugal. Further research in this area could explore detailed data usage and model refinement to advance forecasting efficacy.

Overall, the findings of this project cast a revealing light on Portugal's energy sector. The challenges in utility service scalability urge a rethinking of network management, crucial for service continuity. The influence of environmental factors on renewable energy adoption, particularly UPACs, underscores the need for policies attuned to regional environmental dynamics. In electric mobility, the diverse determinants of EV infrastructure deployment advocate for holistic urban planning approaches. The superiority of *SARIMAX* and hybrid models in forecasting signals a shift towards more accurate, data-driven energy management strategies. Together, these insights form a vital blueprint for developing a more sustainable and efficient energy landscape in Portugal.

Future research on *Integrated Data-Driven Optimization for Sustainable Energy Distribution Networks in Portugal* could explore several avenues. One area involves deepening the understanding of data integration techniques across different energy sources and distribution networks. Another research direction could focus on the application of advanced ML and AI (Artificial Intelligence) algorithms to predict and manage energy demand and supply more efficiently. Additionally, investigating the impact of smart grid technologies on enhancing network resilience and efficiency would be valuable. These studies could

## Group Part

significantly contribute to optimizing Portugal's sustainable energy distribution, aligning with global sustainability goals.

## **6. Rita Oliveira - Electric Mobility and Grid Development**

### **6.1 Introduction**

The transition towards sustainable mobility marks a pivotal moment in the transportation market, mainly through the growing emphasis on *Electric Vehicles* (EVs). Central to this transformation is the development of an accessible and reliable EV charging infrastructure, which promises to address both environmental concerns and the practical needs of EV users.

This paper explores the factors influencing the expansion of EV charging points, known as *Charging Point Equipments* (CPEs), within the context of Portuguese municipalities. The research is conducted by a principal question: To what extent do economic, demographic, and environmental factors correlate with the expansion of EV charging infrastructure, as evidenced by the density of CPEs and the frequency of new grid connections in these regions?

Various complex interaction components are investigated to understand this problem's complexities better. According to the hypothesis, greater population densities are anticipated to correspond with a higher density of CPEs. This shows that urbanisation and the adoption of electric mobility are inseparably connected. Through meticulous research, we investigate how economic, demographic, and environmental considerations correlate with deploying these charging infrastructures.

The research reveals subtle correlations between these factors, suggesting that, while metropolitan areas have a larger concentration of CPEs, there are significant variations caused by municipal economic and environmental policies.

A comprehensive literature review follows a methodology framework chapter that addresses the methodological approach, including data selection and model adoption. The data analysis, results, and discussion are then given, followed by an interpretative appraisal of the findings, contextualization within the hypothesis framework, and discussion regarding their profound implications.

## **6.2 Literature Review**

In the evolving landscape of Electric Mobility, the deployment of EV, CPEs and the associated infrastructure plays a crucial role. This thesis's core is developing and distributing EV charging infrastructure, a crucial aspect of the transition to electric mobility.

Falchetta and Noussan (2021) thoroughly analyze the EV charging network in Europe, focusing on accessibility and deployment trends. They emphasize the importance of considering environmental, economic, and user acceptance factors in developing charging infrastructure. Their work is pivotal in understanding how infrastructure growth may correlate with network connection requests and influence broader electric mobility adoption.

Further, the study by Javid et al. (2019) in California delves into the environmental and economic impacts of expanding public EV charging infrastructure. Their findings demonstrate a significant correlation between charging infrastructure and BEV adoption rates, offering insights into how infrastructure development might impact grid connection demands.

The technological aspects of EV charging and grid integration, alongside consumer behavior, are crucial in understanding the influence of CPE characteristics on electric mobility adoption. The comprehensive review by Das et al. (2020) highlights the significance of technical features, including voltage levels, in adopting and using electric mobility. Simultaneously, the study by White et al. (2022) investigates consumer behavior about public charging station density in the U.S., providing insights into subjective norms and range anxiety. These aspects enrich the understanding of how CPE characteristics can shape consumer preferences and adoption trends.

A comparative analysis of EV adoption strategies, as showcased in the study by Mersky et al. (2016) in Norway, sheds light on various factors influencing BEV sales, such as infrastructure accessibility and regional income. This comparative perspective is critical for understanding how CPE availability impacts electric mobility adoption in different locations.

Additionally, Haidar and Aguilar Rojas's (2022) research on deploying public charging infrastructure in France offers a nuanced view of how different incentives influence BEV and PHEV sales. They identify subjective norms favoring BEVs as a significant influence, whereas concerns like range anxiety play a lesser role, underscoring the importance of CPE characteristics in shaping electric mobility adoption.

Further integrating EVs into the clean energy transition, Wang et al. (2023) propose a model for coordinating the expansion of EV charging infrastructure with renewable energy sources in power grids. Their exploration into EV charging loads and their interaction with renewable energy sources underscores the importance of integrating electric mobility into broader sustainability goals.

This literature review provides a comprehensive and integrated understanding of the factors influencing the development and effectiveness of electric mobility infrastructure. It highlights the interplay between CPE availability, grid development, and EV adoption, offering a cohesive backdrop for the thesis's research questions and hypotheses. This synthesis of scholarly insights sets the foundation for the empirical investigation and analysis in the subsequent parts of the thesis.

### **6.3 Research Question & Hypothesis**

Building on the literature's findings, this research attempts to address this void by presenting the following fundamental question:

- **Research Question:** *“To what extent do economic, populational, and environmental factors correlate with the expansion of EV charging infrastructure, as indicated by the density of CPEs and the frequency of new grid connections in Portuguese municipalities?”*

1. **Hypothesis 1:** Higher population density in Portuguese municipalities is positively associated with greater density of CPEs.

2. **Hypothesis 2:** Higher income levels are positively correlated with higher CPEs density in different municipalities, significantly affecting EV adoption.
3. **Hypothesis 3:** Cities with higher population densities and ZER urban policies in Portugal are likely to have a greater density of CPEs due to higher demand, unlike cities with lower densities and more lenient zoning rules.
4. **Hypothesis 4:** The voltage level of electrical installations is associated with the inclusion of electric mobility.

## 6.4 Methods

### 6.4.1 Data Selection

This research utilized secondary and archival data to investigate CPEs in Portugal, primarily sourced from *Mobi-E*, the national EV charging network. This primary dataset included crucial information like *Postal Code*, *City*, and *Address*, enabling detailed spatial distribution analysis of CPEs. To enhance the geographical accuracy of this data, the Google Cloud Console's *Geocoding API* was employed, converting textual addresses into precise geographical coordinates crucial for visualizing and analyzing CPE distribution patterns.

Additionally, to provide a broader socio-economic perspective, data from *PorData* was incorporated, offering key statistics on population density and average income across Portuguese municipalities. This data helps to understand the electric mobility infrastructure's demand and distribution in different regions, considering historical and geographical factors.

Furthermore, the study also included data from E-Redes, Portugal's leading electricity distributor. This dataset, titled "*Number of Low Voltage Delivery Points with Collection of Load Diagrams*" provided detailed information on the voltage level, installation type, and aspects of electric mobility. This high-quality data from a reputable source added significant value to the study, although its limited historical scope restricted the potential for time series analysis. The combination of these diverse data sources provided a comprehensive view of Portugal's electric mobility infrastructure and socio-economic context.

### 6.4.2 Model Selection

This study employs multifaceted research to investigate the relationship between variables. Several statistical analyses were utilized to address the research question and test the hypothesis effectively, each selected for its suitability to the nature of the data and the specific hypothesis objective.

#### 6.4.2.1 Correlation Analysis

It was an appropriate initial step to quantify the strength of the relationship between *population density*, *income levels*, and *CPE density*. So, two statistical methods were employed:

1. ***Pearson Correlation*** was chosen due to its sensitivity to linear relationships. The coefficient was calculated to determine the degree of linear association between the continuous variables – *population density*, *income levels*, and *CPE density*. The complementary p-value provided a significant test for the correlation, indicating whether the observed association could be attributed to casual.
2. ***Spearman's Rank Correlation***, after recognizing that the data might not meet Pearson's correlation assumptions (such as the presence of outliers or non-normality), Spearman's rank correlation was used as a non-parametric measure that assesses monotone relationships, which is less sensitive to the exact numerical values and more to the rank order of the data points. It was particularly relevant for the analysis as it could validate the *Pearson* results, mitigate the impact of outliers, and provide additional robustness to the findings.

#### 6.4.2.2 Clustering Analysis

K-Means clustering was used to identify patterns in municipalities based on *population* and *CPE density*. This approach helped discover patterns, such as urbanisation or socio-economic factors, by segmenting municipalities into groups with similar characteristics. This segmentation provided strategic insights.

#### **6.4.2.3 Regression Analysis**

**OLS Regression** was employed to model the relationship between population density, average income per city, and CPE density. The addition of a constant ensured accurate model estimations.

**Linear Regression Model** was applied to predict CPE density from urban characteristics. The model was evaluated using the MSE and R-Squared metrics. MSE assessed the model's accuracy, while R-Squared indicated its explanatory power. This *Linear Regression* analysis aimed to predict CPE density from urban characteristics, interpreting the impact of each variable on the outcome.

#### **6.4.2.4 Spatial Distribution Analysis**

A visual representation was conducted using a heat map to subject the spatial distribution of CPEs. This method was chosen because it allows the geographical plotting of CPE densities across cities while providing an intuitive visual representation of data concentration and potential hotspots, which is vital for identifying patterns that might be related to urban population density and policy zones.

#### **6.4.2.5 Chi-Squared Test Analysis**

The *chi-square test* was nominated because it is the standard procedure for testing the relationship between two categorical variables. The contingency table provides the practical frequencies of electric mobility inclusion across different *voltage levels*, allowing the *chi-square test* to assess whether the observed distribution differs from expected under the assumption of independence.

#### **6.4.2.6 Cross-Validation Analysis**

A K-fold cross-validation was performed to validate the robustness of the *chi-square test results*. Using ten folds (k=10) is a common practice that provides a good equilibrium between bias and variance. It ensures that every observation from the

original dataset has a performance change in the training and test set, which is crucial to prevent overfitting and ensure the stability of the test results.

#### **6.4.2.7 Logistic Regression Analysis**

*Logistic Regression* was chosen for its ability to handle a binary dependent variable, in this case, to estimate the probability of electric mobility inclusion based on voltage level categories (presence or absence of electric mobility). It is beneficial for understanding the strength of the relationship between the dependent and independent variables. Hyperparameter tuning used *GridSearchCV* to optimize the logistic regression model's parameters. This technique systematically explores a range of hyperparameters to find the finest combination that maximizes the model's accuracy, ensuring that the model's performance is as high as possible. *Sensitivity analysis* by changing the test size was led to evaluate the consistency of the logistic regression model's accuracy. This approach helps to comprehend the impact of the dataset division on the model performance, ensuring that the conclusions drawn from the model are not dependent on a precise part of the data.

### **6.5 Data Analysis**

#### **6.5.1 Data Description**

*Dataset 1* contains a range of variables that provide detailed geographical and demographic data on Portuguese municipalities and includes a comprehensive structure of 308 rows (which corresponds to the number of municipalities in Portugal) and eight columns. The variables included in the dataset are as follows:

- **City:** Catalogs all cities within Portugal, offering a detailed breakdown of urban centers.
- **Latitude and Longitude:** which provides precise locational coordinates for each city.
- **CPE\_Count:** Quantifies the number of CPEs present in each city.
- **Km2:** Indicates the land area of each city in square kilometers.
- **Density of CPE per City:** Number of CPEs per square kilometer of each city.
- **Population Density per City:** Number of inhabitants per square kilometer.

*Dataset 2* is an enriched dataset of *Dataset 1* that now includes a new variable that provides a deeper investigation by incorporating economic factors into the analysis.

- **Average Income per City:** Average income metrics for each Portuguese city.

*Dataset 3* has 55997 rows and 10 columns, providing a complete overview of EV charging stations in Portuguese municipalities. The following variables are included in the dataset:

- **Parish Code:** Unique identifier for each parish within the municipality.
- **Date:** The time point when the data was recorded, formatted as “YYYY-MM”
- **City:** The specific city where the charging point is located.
- **Municipality:** The broader municipal area encompassing the charging point
- **Inc Electric Mobility:** Indicates if the area is included in electric mobility initiatives.
- **CPEs:** Total number of CPEs in the area.
- **Number of CPEs with collected DC:** Count of CPEs that provide Direct Current (DC) charging.
- **Type of Installation:** Classifies the CPEs as Domestic, Non-Domestic, or Other.
- **Voltage Level:** Which voltage level does the CPE operate, such as Low, Medium, or High Voltage

### 6.5.2 Exploratory Data Analysis

An exploratory data analysis (EDA) was done before beginning the modelling phase to visualise the data distribution, identify outliers, and uncover underlying patterns or trends.

*Table 1. Descriptive Statistics for Dataset 1.*

	<i>CPE_Count</i>	<i>Km2</i>	<i>Density of CPE per City</i>	<i>Population Density per City</i>
<b>Count</b>	308	308	308	308
<b>Mean</b>	16.317308	313.127981	0.112332	6.715762
<b>Standard Deviation</b>	24.379949	294.173993	0.221275	34.707289
<b>Minimum</b>	1	7.940000	0.001412	0.004934
<b>25%</b>	3	124.7775	0.013615	0.151712
<b>50%</b>	8	217.045	0.028426	0.492450
<b>75%</b>	18	407.3975	0.100811	2.254154
<b>Maximum</b>	198	1720.6	1.766190	365.377834

Appendix S displays the distribution of CPEs density across the Metropolitan Area of Lisbon, using a color gradient to represent varying densities. Warmer colors indicate areas with higher CPEs density. And Appendix T the same but in Metropolitan Area of Porto.

According to Appendix U, while increasing population density would theoretically correspond to higher CPE densities because of increased demand for EVs, the actual situation may be altered by other variables.

Appendix V depicts the variation in city sizes and illustrates that, although some cities have small regions, others have large territories. This heterogeneity can impact urban planning, resource allocation, and infrastructure construction.

Appendix W illustrates a plot that shows the relationship between various variables in the dataset; each plot indicates a pairwise relationship. Histograms describe the distribution of a single variable along the diagonal. One may see how one variable is connected to another by looking at the pair plots off the diagonal. For example, the plot where *Latitude* intersects with *Longitude* displays the geographical distribution of the data points. In contrast, the intersection of *CPE Count* with *Population Density per City* may suggest whether there is a link between the availability of CPEs and population density. These visualisations are very effective for identifying trends, correlations, and outliers in multidimensional data.

Positive correlations are represented in Appendix X by warmer colors (red), implying that if one variable grows, so does the other. Cooler colors (blues) reflect negative correlations, meaning that if one variable grows, the other tends to decrease. Darker colors (black) represent no or weak correlation, implying little to no linear relationship between the variables.

A scatter plot with a regression line was produced in Appendix Y to visually analyze the link between population density and CPE density. The figure shows a positive tendency, with cities with greater population densities having higher CPE densities. The extent and direction of this link are shown by the regression line and the shaded region surrounding it, which indicates the confidence interval.

Appendix Z depicts the top 20 cities regarding CPE density per square kilometer. It shows CPE density in descending order, with the highest density on the left side of the figure. This visualisation assists in rapidly identifying cities with the largest concentration of CPEs about their surrounding area.

A K-Means clustering analysis was carried out with the assumption of three clusters. Appendix AA depicts the distribution of cities by population and CPE density, with each cluster represented by a distinct color. Each cluster's center is indicated with a black dot. The clusters may point to trends in the relationship between population density and CPE density across distinct groupings of cities.

The histogram in Appendix BB depicts the frequency distribution of average income among cities. Most cities have an average income of roughly 900 to 1100, with a peak around 1000, suggesting that this is the most frequent average income range among the cities surveyed. The right tail is lengthy, indicating that fewer cities have much higher average earnings, but the frequency of cities with higher incomes reduces significantly following the peak.

Appendix CC depicts the relationship between *Average City Income* and *CPE Density*. The cluster of cities with the lowest average income has the lowest CPE density. As the average income rises, CPE density varies, with a few places having a significantly high density.

In Appendix DD the row and column of *Average Income per City* shows a moderate positive correlation (0.5) with CPE Count suggesting that cities with higher average income may have higher count of CPEs. It has a low positive correlation with *Population Density per City* (0.14) and almost no correlation with *Latitude*, *Longitude*, and the size of the city, *Km2*.

The bar chart in Appendix EE shows a clear ranking of cities depending on the number of CPEs. Certain cities, such as "*União das freguesias de Cascais e Estoril*" clearly have more CPEs than others. This higher number might be due to more electric car use, progressive urban design, or increased economic resources given to electric mobility infrastructure in certain

countries. The large diversity in the number of CPEs among the top cities shows that different objectives or phases of development in the spread of electric mobility solutions may exist.

Appendix FF compares the average number of CPEs in places where electric mobility is available vs those where it is not. The stark disparity, with a significantly higher average in regions with electric mobility inclusion, highlights the potential usefulness of policies and encourage the use of EV. This high number shows that taking preemptive steps in electric transportation might result in a more resilient infrastructure to accommodate such technology.

### **6.5.3 Results Report**

The *Pearson Correlation Analysis* for Hypothesis 1 produced a coefficient of 0.79 and a p-value  $< 0.001$ . This implies a high, positive association between *Population Density per City* and *CPE density per City*, which is statistically significant since the p-value is considerably below the 0.05 threshold.

The *Spearman's Rank Correlation* analysis resulted in a coefficient of 0.69 and a low p-value  $< 0.001$ . This also shows a substantial positive correlation between *Population Density per City* and *Density of CPE per City* when ranked, confirming the presence of a predictive relationship. A K-Means clustering analysis was carried out with the assumption of three clusters. Appendix AA depicts the distribution of cities by population and CPEs Density, with each cluster represented by a distinct color. Each cluster's center is indicated with a black dot. The clusters may point to trends in the relationship between *Population Density* and *CPE Density* across distinct groups of cities.

For Hypothesis 2, the *Pearson Correlation Coefficient* of 0.3 suggests a low to moderate positive linear relationship between average income and density of CPEs per City. The p-value  $< 0.001$  indicates that this correlation is statistically significant, meaning there is a very low probability that this correlation is due to random chance. While the *Spearman's Rank Correlation Coefficient* of 0.5 suggests a moderate positive monotonic relationship

between the two variables. The p-value < 0.001 further confirms the statistical significance of this relationship.

The conducted *Regression Analysis* aimed to ascertain the determinants of CPEs density in urban areas. The dependent variable *Density of CPE per City* was regressed against two independent variables: *Population Density per City* and *Average Income per City*.

Table 2. OLS Regression Results - Density of CPE per City.

OLS Regression Results	
Dependent Variable	Density of CPE per City
Method	Least Squared
Feature	Coefficient
R-squared	0.673
Adjusted R-squared	0.670

The resultant model exhibited an R-Squared value of 0.673, implying that approximately 67.3% of the variance in CPE density is explicable by the included independent variables. Accounting for the number of predictors, the Adjusted R-Squared of 0.670 offers a modestly cautious estimate.

Table 3. Regression Analysis Summary.

	Const	Population Density per City	Average Income per City
Coefficient	-0.1722	0.0049	0.0002
Standard Deviation Error	0.053	0.000	5.05e-05
t	-3.247	19.139	4.790
P> t	0.001	0.000	0.000
[0.025	-0.277	0.004	0.000
0.975]	-0.068	0.005	0.000

Statistical testing found that the *Population Density per City* coefficient of 0.005 and the *Average Income per City* coefficient of 0.0002 positively influence CPE Density, with the population density having a larger impact. The significance of these predictors was confirmed by their p-values, which were found to be less than 0.001, substantiating the hypothesis that both variables are critical in predicting CPE Density.

For Hypothesis 3, *Heat Mapping* visualisation highlighted geographical tendencies in the distribution of CPEs, revealed a non-uniform distribution, with larger concentrations seen

in various metropolitan areas. The dataset simplified the examination of CPE Density about urban planning constraints, particularly in cities like Lisbon and Porto, which have implemented street restrictions. An analysis function was employed to categorize each CPE location as either within a restricted zone or not. The comparison of average CPE Densities revealed a higher mean density within restricted zones (0.359) compared to non-restricted zones (0.279), suggesting that policy interventions may impact the spatial arrangement of CPEs.

In the case of Hypothesis 4, to cross-tabulate the counts of *incl Electric Mobility* against *Voltage Level categories*, a *Contingency Table* was built:

Table 4. Contingency Table.

	Included in Electric Mobility	Not Included in Electric Mobility
<b>Low Voltage Normal</b>	5 554	32 483
<b>Low Voltage Special</b>	2 267	7 289
<b>Very High, High and Medium Voltage</b>	1 403	7 001

The *Chi-Square Test of Independence* was then used to establish a statistically significant relationship between the two variables, the output of 462.2 suggests a significant relationship between the incorporation of electric mobility and electrical installation voltage levels. The extraordinarily low p-value  $< 0.001$  strongly contradicts the null hypothesis of independence, implying that electric mobility is not distributed equally across different voltage levels.

A *Cross-Validation* was conducted using K-Fold (with  $k=10$ ) to validate the robustness of the chi-square test results. Each fold produced a significant chi-square statistic, varying from 18.165 to 76.954, with corresponding p-values indicating statistical significance in each case. This consistency across folds reaffirms the initial findings from the *Chi-Square Test*.

*Logistic Regression Analysis* was used to quantify the strength of the association between *Voltage Level* (as independent variables) and *incl Electric Mobility* (as the dependent variable).

Table 5. Classification Performance Metrics.

	Precision	Recall	F1-Score
<b>Class 0 (Not Included)</b>	0.83	1.00	0.91
<b>Class 1 (Included)</b>	0.00	0.00	0.00
<b>Accuracy</b>	-	-	0.83
<b>Macro Average</b>	0.42	0.5	0.45
<b>Weighted Average</b>	0.69	0.83	0.76

The model predicted electric mobility inclusion with 83% accuracy but did not predict the “Included” class well, as indicated by the low recall for that class. The *Logistic Regression* coefficients were transformed into odds ratios, revealing that locations with *Low Voltage Special* are about 1.8 times more likely to include electric mobility than the baseline *Low Voltage Normal*. Locations with *Very High, High, and Medium Voltages* are about 1.18 times more likely to include electric mobility.

*Sensitivity Analysis* examined the effect of varying the test size on model accuracy. The accuracy remained consistent across different test sizes, ranging from approximately 95.46% to 95.57%, indicating that the model's predictions are robust to the proportion of data allocated for testing.

Table 6. Model Accuracy at Varying Threshold Levels.

	Accuracy
<b>0.2</b>	0.955625
<b>0.3</b>	0.954821
<b>0.4</b>	0.954596
<b>0.5</b>	0.954713

Hyperparameter tuning was performed for the *Logistic Regression* model using *Grid SearchCV*, varying the regularization strength and type of penalty. The best model achieved a n accuracy score through cross-validation, and the best parameters were identified for the logistic regression model, which can be used to refine predictions. Fitting 5 folds for each 10 candidates, totaling 50 fits. The best C (which controls the strength of regularization) is 0.01 with a L1 penalty that gives the best cross-validating accuracy score of approximately 95.5%.

#### **6.5.4 Predictive Models**

In Hypothesis 3 the predictive variables included *Km2* for city size, and *Population Density per City* for the number of individuals per square kilometer. These features (x) and the target variable (y), *Density of CPE per City*, were divided into training (80%) and testing (20%) cohorts. This split was performed using a random seed to ensure the replicability of results.

The training data was fitted with a *Linear Regression Model*. Its performance was compared to a test set suspended throughout the training phase to provide an unbiased assessment of the model's predictive capability. The MSE was calculated with a value of 0.035. The MSE quantifies the average of the squares of the errors, indicating the difference between the observed actual outcomes and the values predicted by the model. And the model produced an R-Squared of 0.475, indicating that approximately 47.54% of the variation in the density of CPEs is explainable by the model's inputs.

#### **6.6 Discussion**

The study's statistical analysis reveals a significant positive correlation between population density and CPEs density in Portuguese municipalities. *Pearson and Spearman correlation* analyses demonstrate this relationship with high positive coefficients, indicating a robust linear and monotone correlation. Specifically, the *Pearson correlation* coefficient of 0.724 suggests a robust linear relationship, while *Spearman's correlation* confirms this with a low p-value, asserting its reliability and resistance to outliers.

The study employs *Clustering Analysis*, likely using a K-Means algorithm, to categorize municipalities into three clusters based on population and CPE density. Cluster 0 encompasses areas with low population and CPE density, Cluster 1 includes fewer cities but with a higher CPE density, and Cluster 2, the smallest group, comprises areas with high population density but varying CPE densities. This clustering suggests deficient CPE distribution in high-density areas and targeted EV infrastructure development in less populated areas.

From a practical standpoint, these findings imply that municipalities with higher population densities are more suitable for investments in CPE infrastructure. Strategically, promoting EV usage should focus on densely populated areas for maximum impact. Theoretically, the research contributes to the literature on innovation diffusion, especially regarding sustainable technologies, by providing empirical evidence of the distribution of EVs and related infrastructure in urban areas. Population density can be a driving factor for technological adoption.

However, the study challenges Hypothesis 2, which posited a positive correlation between higher income levels and CPE density. Regression analysis, factoring in population density and average income per municipality, identifies population density as a significant predictor for CPE density but not average income. The model's R-Squared value of 0.673 indicates that these factors explain a substantial portion of CPE density variability, but income levels do not strongly predict CPE densities.

This finding contradicts Hypothesis 2 and suggests that wealth does not directly correlate with the growth of CPE infrastructure when population density is considered.

Additionally, diagnostic tests of the regression model reveal a non-normal residual distribution, suggesting a more complex relationship between income and CPE density than what a linear model captures. Although scatter plot analysis shows a slight positive trend between average income and CPE density, the connection is not strong and is likely influenced by other factors or outliers.

The study also examines the role of policy, particularly in cities with free EV parking policies, which tend to have higher average incomes. However, these policies do not statistically enhance CPE density, indicating that income levels and incentives like free parking are more complex predictors.

Hypothesis 3 suggests that cities with higher population densities and the implementation of ZER urban policies in Portugal will likely exhibit a greater density of CPEs. The analysis

shows a weaker correlation between ZER policy zones and CPE density, indicating that ZER policies alone are not decisive in CPE infrastructure proliferation. However, restricted zones within cities, part of ZER policies, exhibit a higher average density of CPEs, supporting the hypothesis that regulatory measures can encourage CPE installation.

Predictive modeling, using city size and population density as predictors, accounts for about 50% of the variance in CPE density, suggesting that while these factors are influential, other variables also play a role in determining CPE density.

Finally, Hypothesis 4 links the voltage level of electrical installations to electric mobility adoption. Statistical tests, including the *Chi-Square Test of Independence*, show a significant association between certain voltage levels and electric mobility inclusion. This is further confirmed by *Logistic Regression* analysis and cross-validation, indicating that low-voltage unique installations are more likely to be associated with electric mobility. This finding has implications for urban development and planning of electric mobility infrastructure, suggesting that focusing on areas with specific voltage levels can facilitate the growth of EVs.

### **6.6.1 Limitations of the Study**

In the limitations section of this thesis, it is essential to acknowledge the constraints encountered during the research process. While there was an aspiration to delve deeper into the study, the limitations posed by the data available proved to be a significant hindrance. Efforts were made to circumvent these limitations by seeking alternative data types, such as *population density*, *CPEs locations*, and *income levels*. However, these attempts were challenged due to the unavailability of comprehensive information and the need for more temporal data. The absence of time-series data impeded the ability to conduct a thorough temporal analysis, which could have provided more nuanced insights into the trends and patterns over time.

Consequently, the findings and conclusions of this study must be contextualized within the scope of these data limitations. The data provided had constraints in terms of its breadth but also its depth and temporal range. This limitation inherently impacts the extent to which

the results can be generalized, and the depth of the conclusions drawn from the analysis. Future research in this area would benefit from access to more comprehensive datasets, including longitudinal data, to overcome these limitations and build upon the groundwork laid by this study.

## **6.7 Conclusion**

The comprehensive analysis presented in this paper provide robust support for the initial hypothesis, affirming that higher population density is indeed a compound for the densification of CPEs in Portuguese municipalities. This correlation, authenticated through various statistical techniques, underscores the critical role of urban density in the strategic placement of EV charging infrastructure.

However, the findings extend beyond mere correlations, enlightening the multifaceted nature of infrastructure deployment, which is also influenced by economic, policy, and technical factors. The study challenges defined concepts regarding income levels, policy incentives, highlighting a more complex reality where these elements interact on complex ways. Urban planning and policy formulation emerge as key elements in this equation, with the strategic implementation of ZER policies and considering electrical grid capabilities proving instrumental. As the world leans into a future of sustainable transportation, the insights gathered from this study offer valuable guidance for the multidimensional approach needed in planning and policymaking for EV infrastructure deployment.

This research contributes to the academic understanding of electric mobility and provides actionable insights for effective urban deployment and the fostering of innovative technologies in the dominion of sustainable transport.

## References

- White, Lee V., André Carrel, Wei Shi, and Nicole D. Sintov. 2022. "Why Are Charging Stations Associated with Electric Vehicle Adoption? Untangling Effects in Three United States Metropolitan Areas." *Energy Research & Social Science* 89 (July): 102663. <https://doi.org/10.1016/j.erss.2022.102663>.
- Sterman, John. 2011. "Electric Vehicle Charging Infrastructure Deployment : Policy Analysis Using a Dynamic Behavioral Spatial Model." 2011. <https://dspace.mit.edu/handle/1721.1/65504>.
- Pagany, R. 2018. "A Review of Spatial Localization Methodologies for the Electric Vehicle Charging Infrastructure." 2018. <https://www.semanticscholar.org/paper/A-review-of-spatial-localization-methodologies-for-Pagany-Camargo/267147812beb17f73158737e9be2ef4addb1369f>.
- Narasipuram, Rajanand Patnaik, and Subbarao Mopidevi. 2021. "A Technological Overview & Design Considerations for Developing Electric Vehicle Charging Stations." *Journal of Energy Storage* 43 (November): 103225. <https://doi.org/10.1016/j.est.2021.103225>.
- Mersky, Avi Chaim, Frances Sprei, Constantine Samaras, and Sean Qian. 2016. "Effectiveness of Incentives on Electric Vehicle Adoption in Norway." *Transportation Research Part D: Transport and Environment* 46 (July): 56–68. <https://doi.org/10.1016/j.trd.2016.03.011>.
- Mortimer, Benedict J., Christopher Hecht, Rafael Goldbeck, Dirk Uwe Sauer, and Rik W. De Doncker. 2022. "Electric Vehicle Public Charging Infrastructure Planning Using Real-

## Group Part

World Charging Data.” *World Electric Vehicle Journal* 13 (6): 94. <https://doi.org/10.3390/wevj13060094>.

Javid, Roxana J., Mahmoud Salari, and Ramina Javid. 2019. “Environmental and Economic Impacts of Expanding Electric Vehicle Public Charging Infrastructure in California’s Counties.” *Transportation Research Part D: Transport and Environment* 77 (December): 320–34. <https://doi.org/10.1016/j.trd.2019.10.017>.

Haidar, Bassem, and Maria Teresa Aguilar Rojas. 2022. “The Relationship between Public Charging Infrastructure Deployment and Other Socio-Economic Factors and Electric Vehicle Adoption in France.” *Research in Transportation Economics* 95 (November): 101208. <https://doi.org/10.1016/j.retrec.2022.101208>.

Gouveia, Clara, David Rua, Filipe Soares, Carlos Moreira, Pedro Matos, and João Lopes. “Development and Implementation of Portuguese Smart Distribution System.” *Electric Power Systems Research* 120 (March 1, 2015): 150–62. <https://doi.org/10.1016/j.epsr.2014.06.004>.

Falchetta, Giacomo, and Michel Noussan. 2021. “Electric Vehicle Charging Network in Europe: An Accessibility and Deployment Trends Analysis.” *Transportation Research Part D: Transport and Environment* 94 (May): 102813. <https://doi.org/10.1016/j.trd.2021.102813>.

Das, Himadry Shekhar, Mohammad Mominur Rahman, Shuhui Li, and Chee Wei Tan. 2020. “Electric Vehicles Standards, Charging Infrastructure, and Impact on Grid Integration:

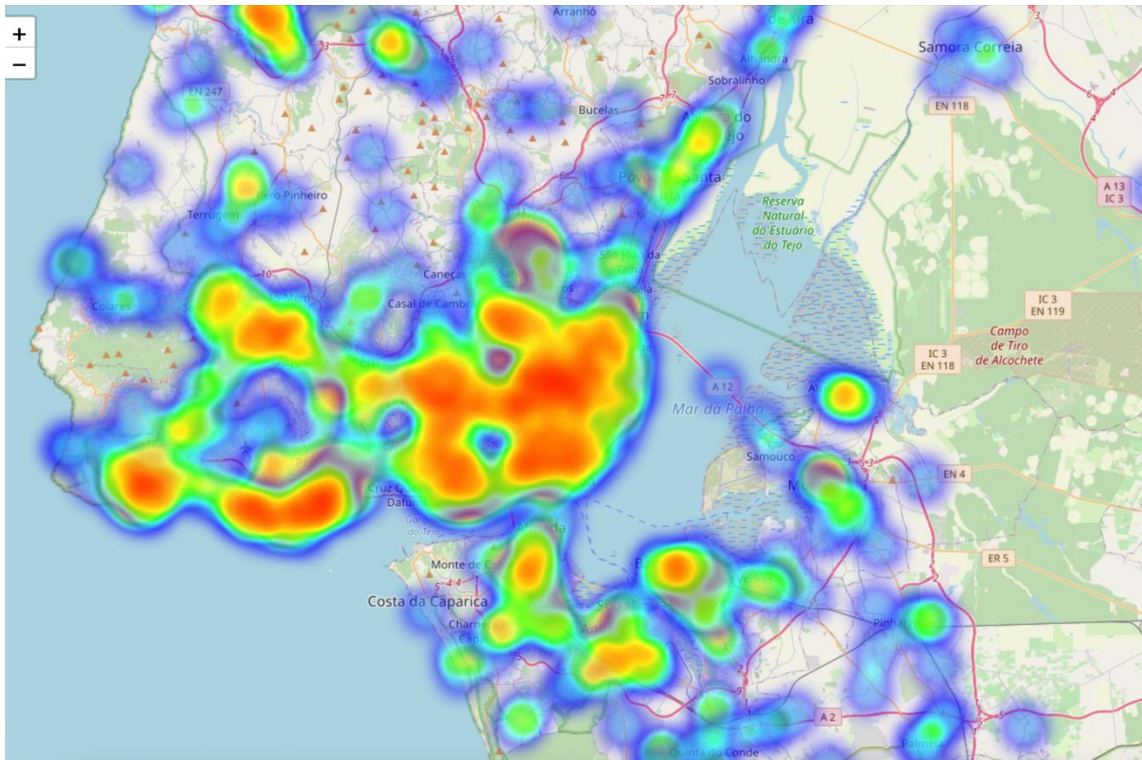
## Group Part

A Technological Review.” *Renewable & Sustainable Energy Reviews* 120 (March): 109618. <https://doi.org/10.1016/j.rser.2019.109618>.

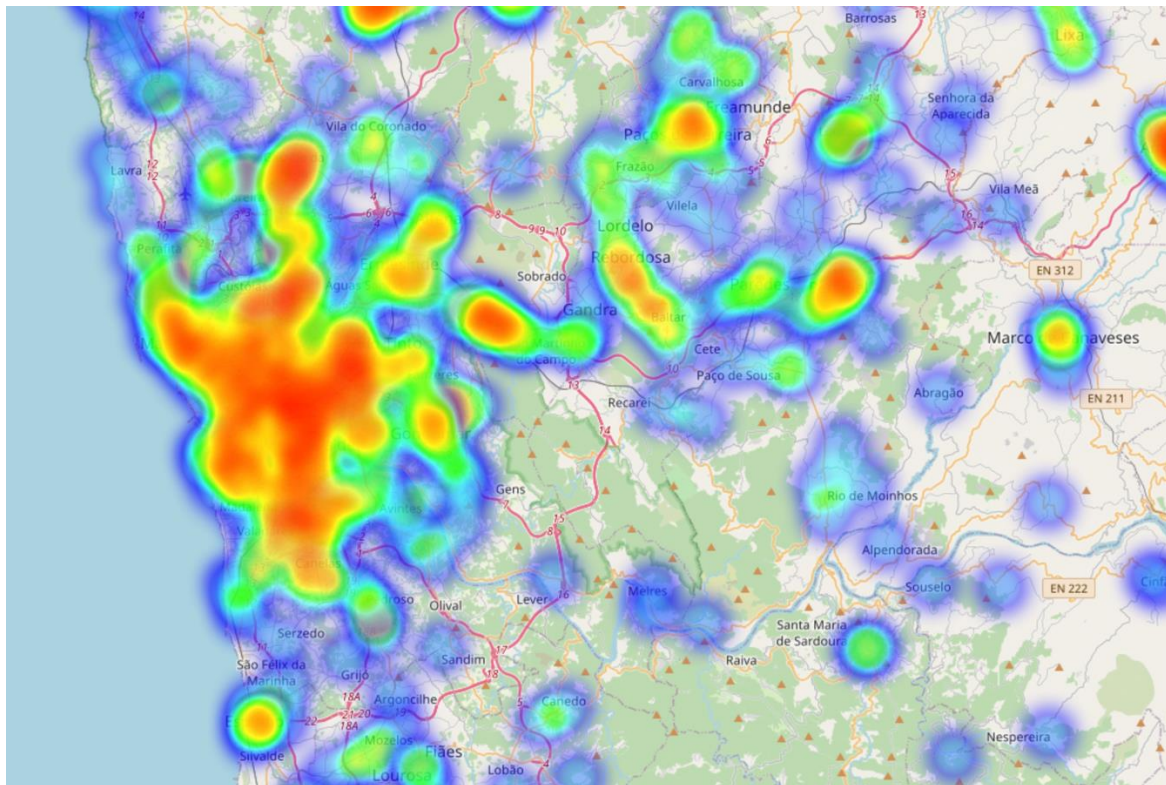
Bo Wang, Payman Dehghanian and Dongbo Zhao.2023. “Coordinated Planning of Electric Vehicle Charging Infrastructure and Renewables in Power Grids.” *IEEE Journals & Magazine*. <https://ieeexplore.ieee.org/abstract/document/10045673>.

Appendix

Appendix A. HeatMap for CPE Density in Lisbon Metropolitan Area.

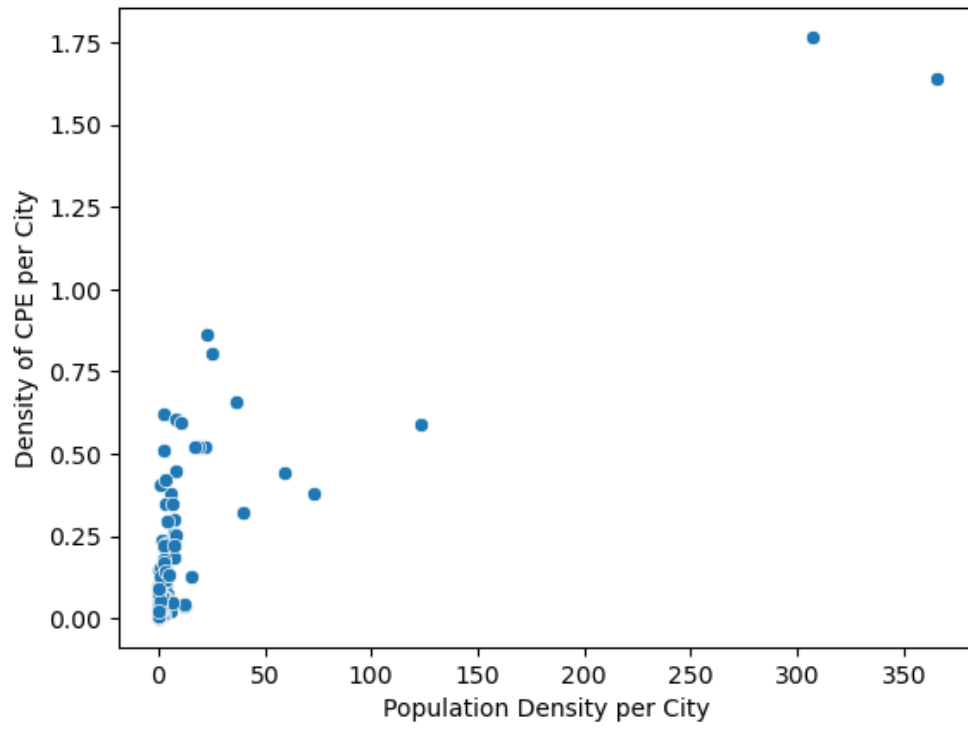


Appendix B. HeatMap for CPE Density in Porto Metropolitan Area.

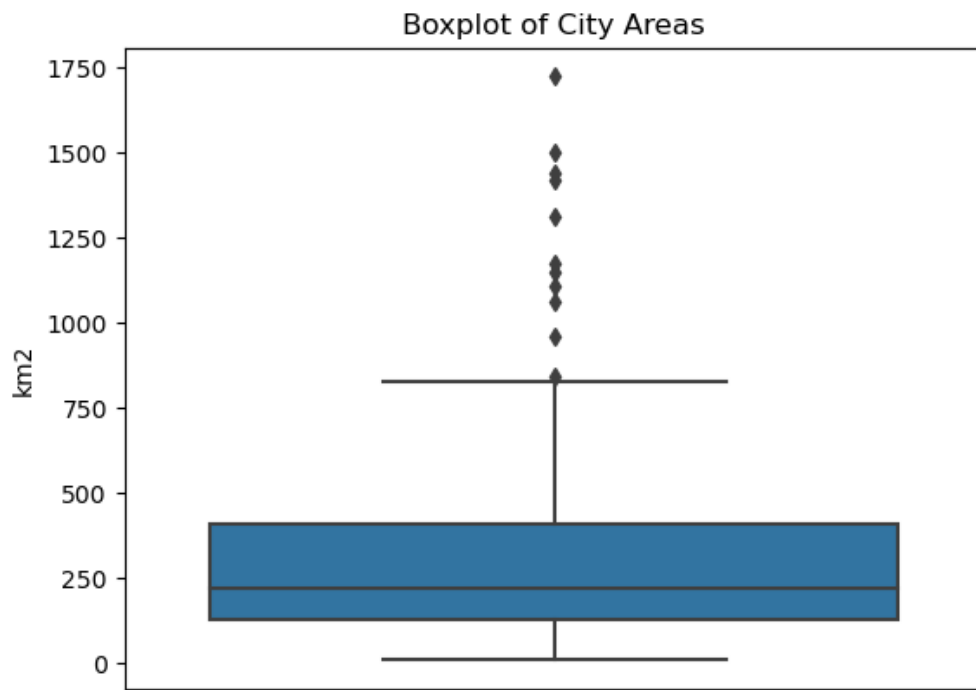


## Group Part

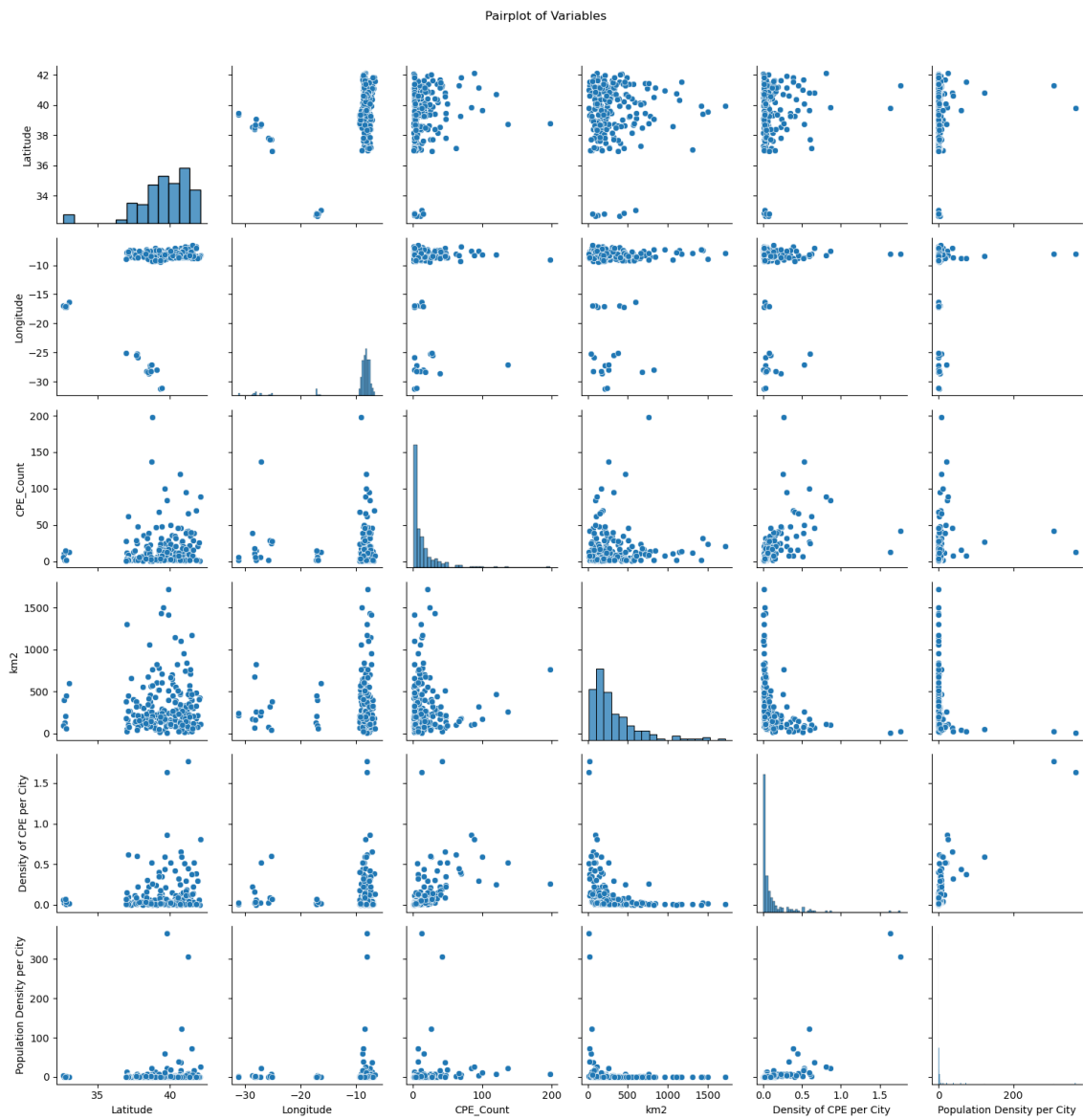
Appendix C. Scatter Plot of CPEs Density versus Population Density.



Appendix D. Box Plot of City Areas.

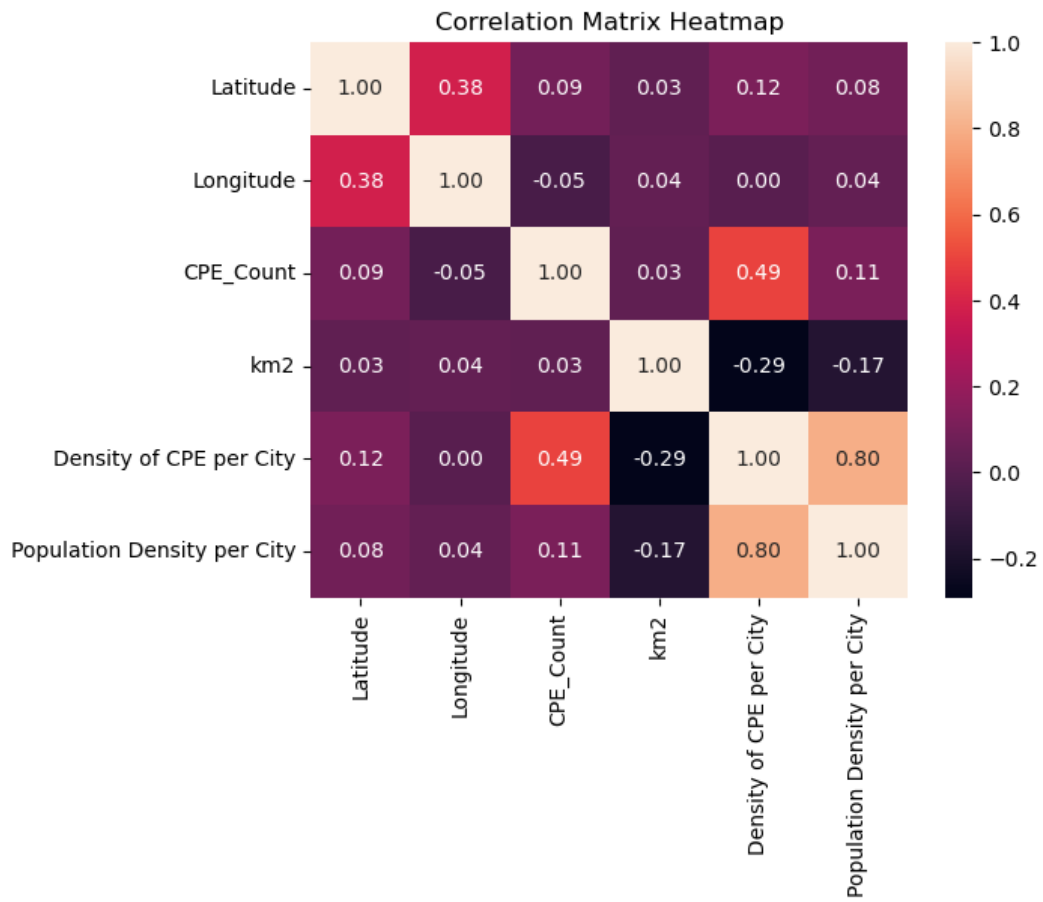


Appendix E. Pair Plot of Variables.

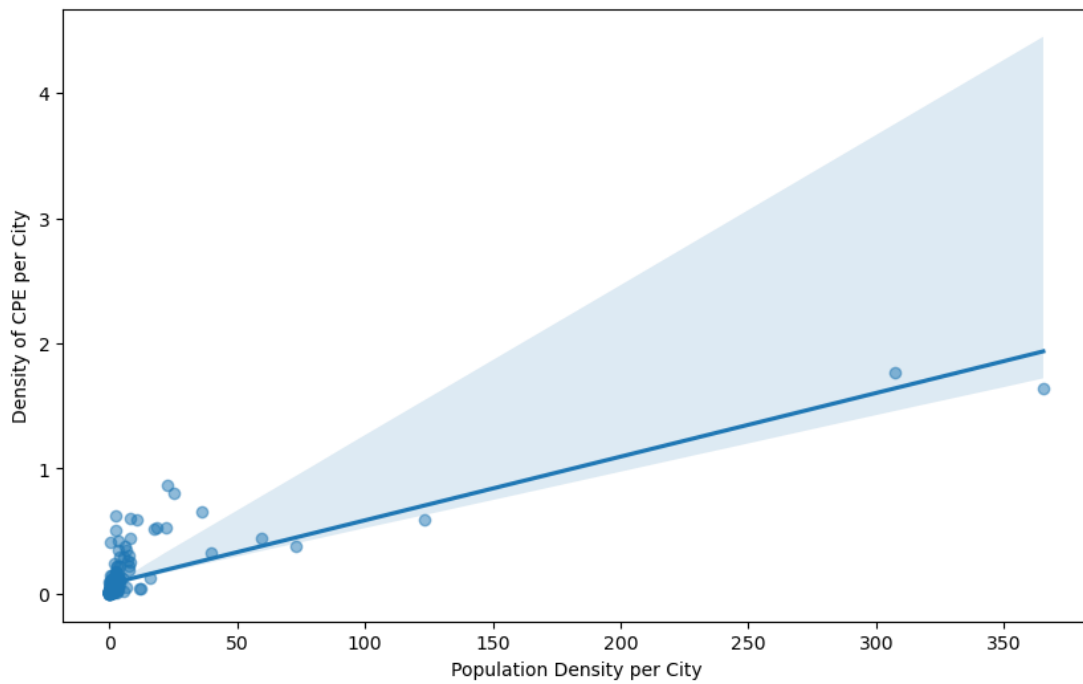


## Group Part

Appendix F. Correlation Matrix Heat Map.

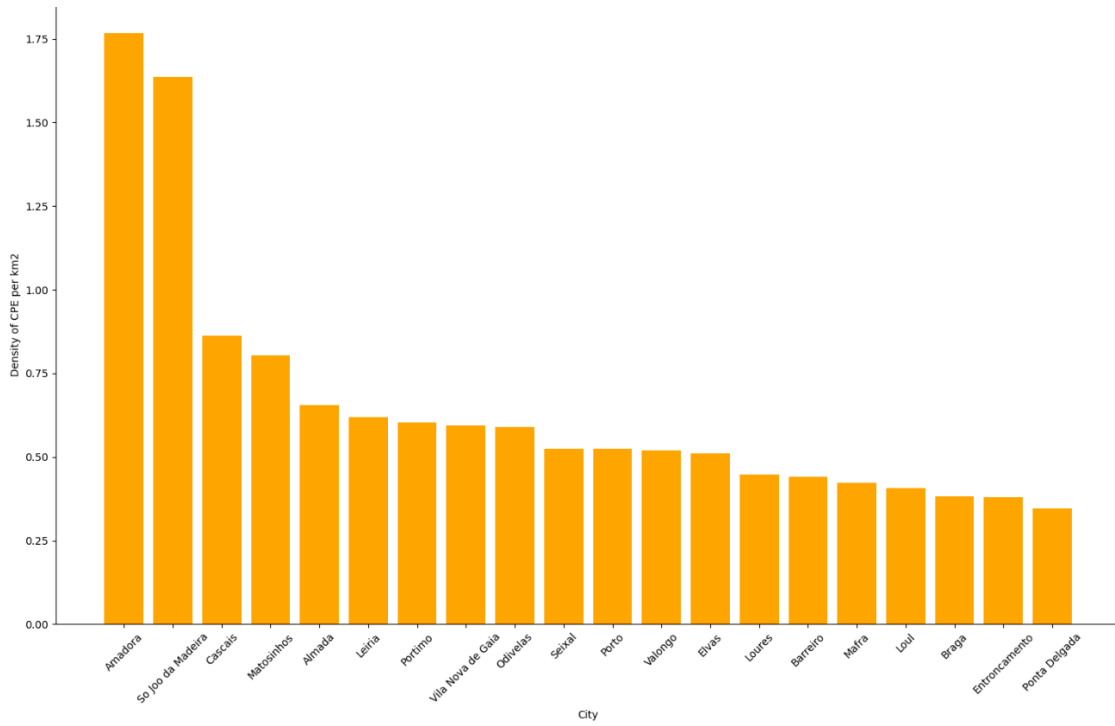


Appendix G. Relationship Between Population Density and CPE Density.

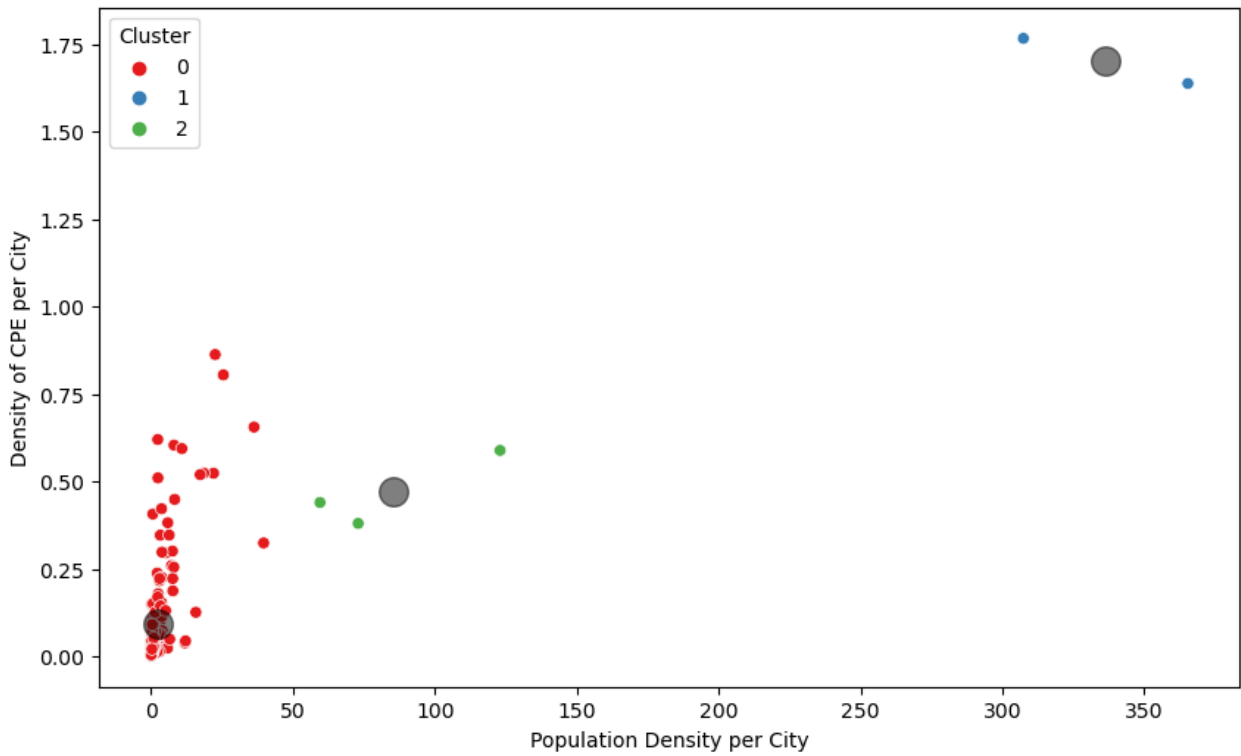


# Group Part

Appendix H. Top 20 Cities by Density of CPE per City.

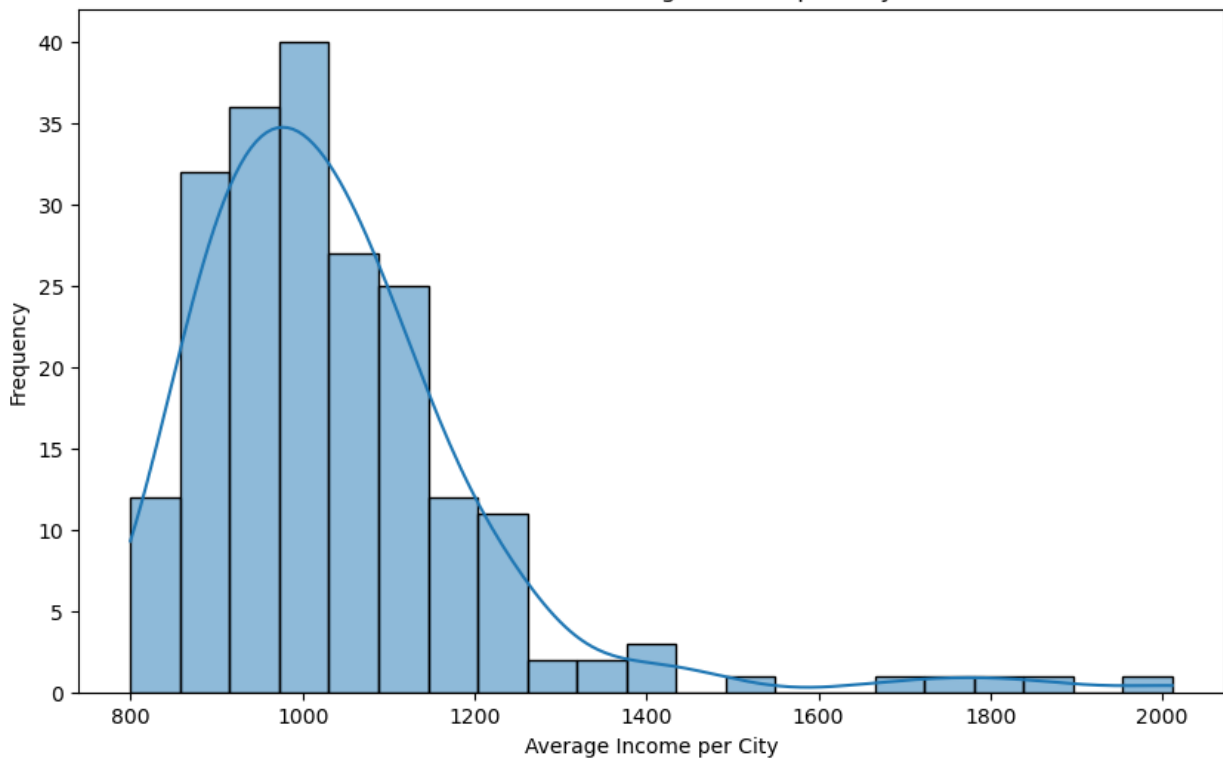


Appendix I. Clustering of Cities by Density of Population and CPE.

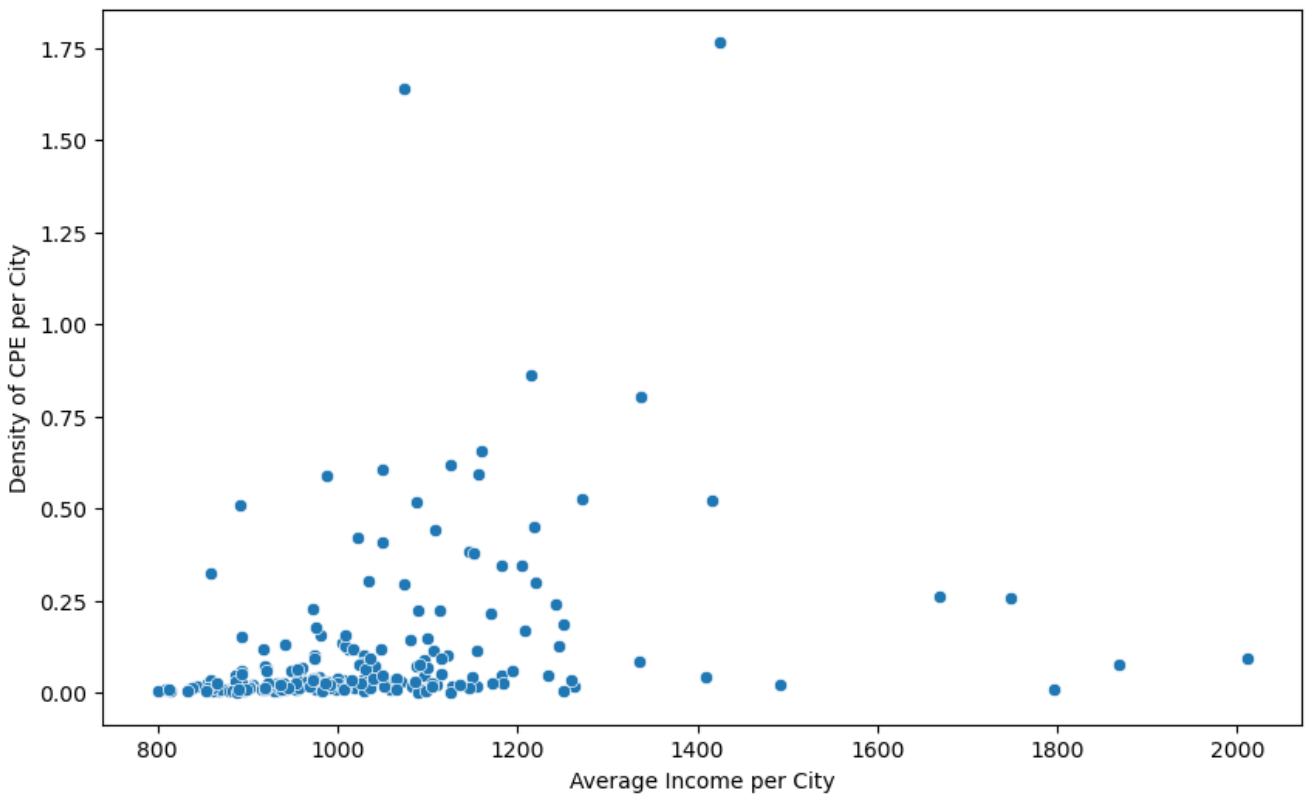


Group Part

Appendix J. Distribution of Average Income per City.

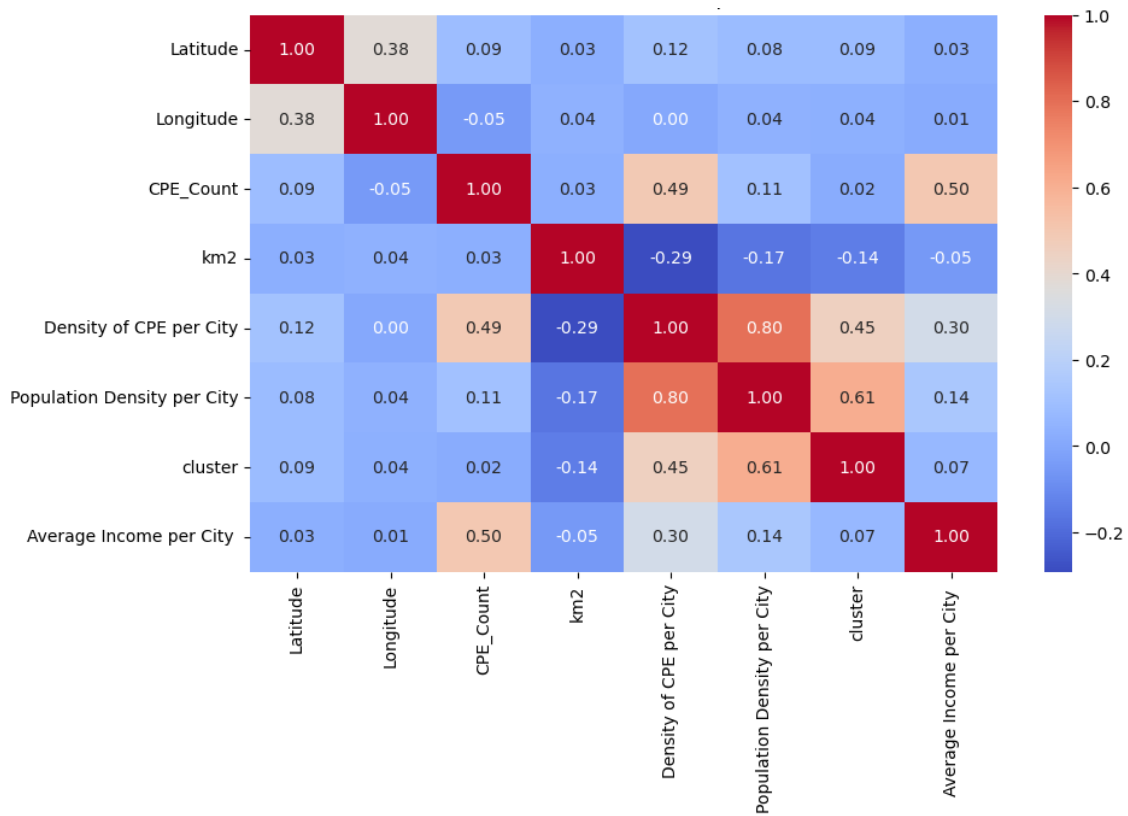


Appendix K. Relationship between Average Income and CPE Density per City.

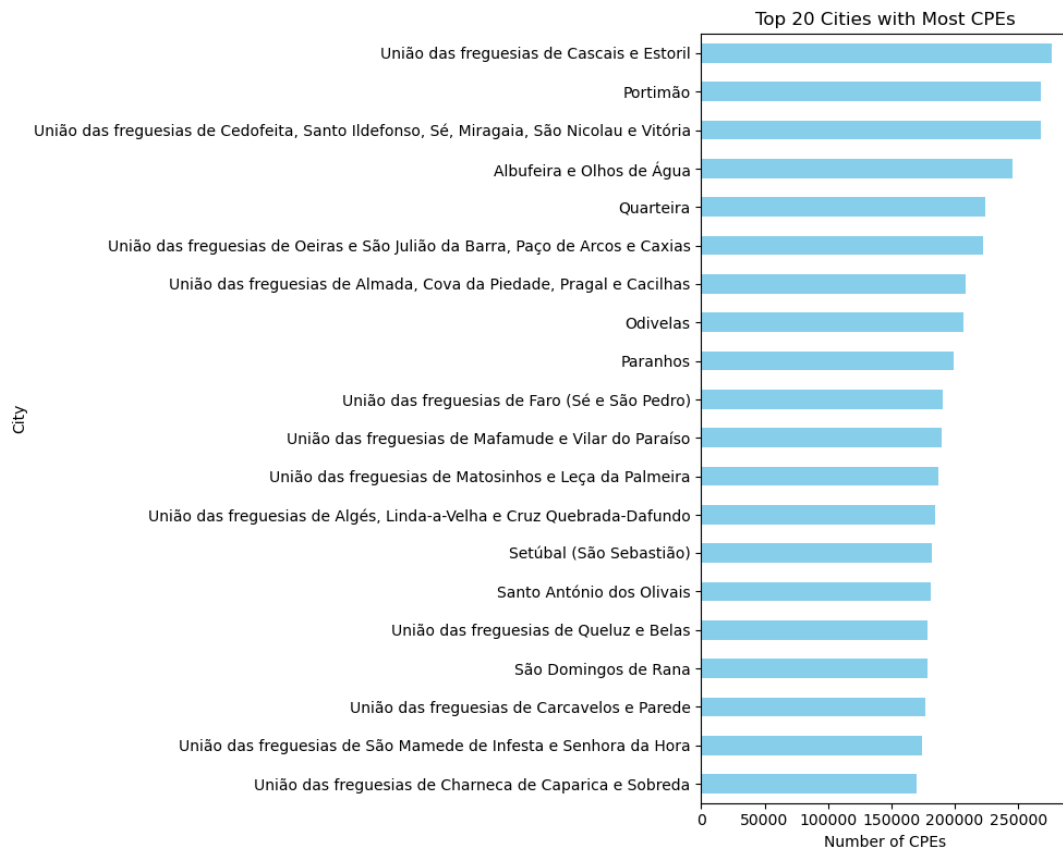


## Group Part

Appendix L. Correlation Heat Map with Average Income per City.



Appendix M. Top 20 Cities with Most CPEs.



## Group Part

*Appendix N. Average CPEs by Electric Mobility Inclusion.*

