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.

SERVICE CONTINUITY IMPACTING REMOTE SERVICES

DUARTE NUNO BRAGA CORREIA

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

List of Figures

List of Tables

Abbreviations

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

Group Part

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

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.

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, 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-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

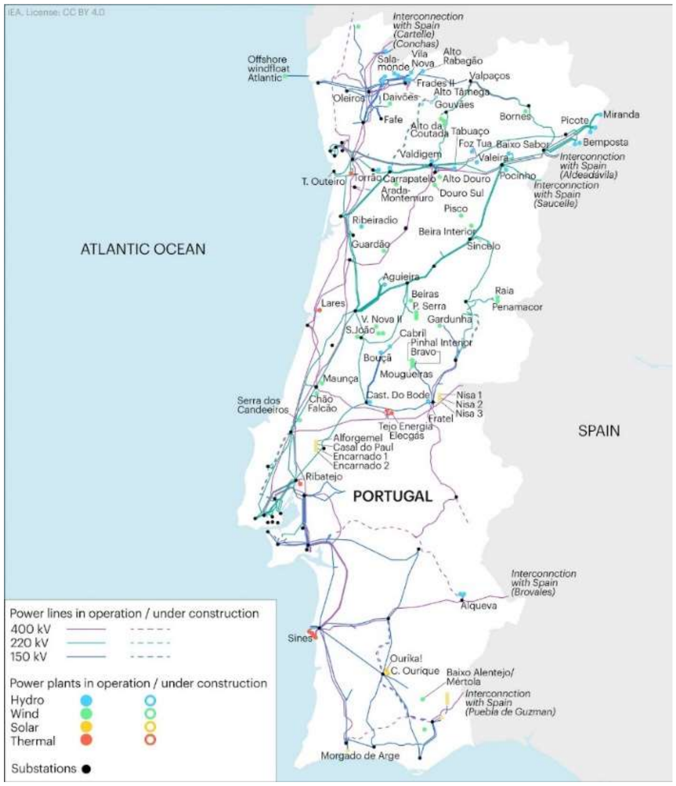


Figure 1. Portugal National Transport Grid Map. Source: REN.

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).

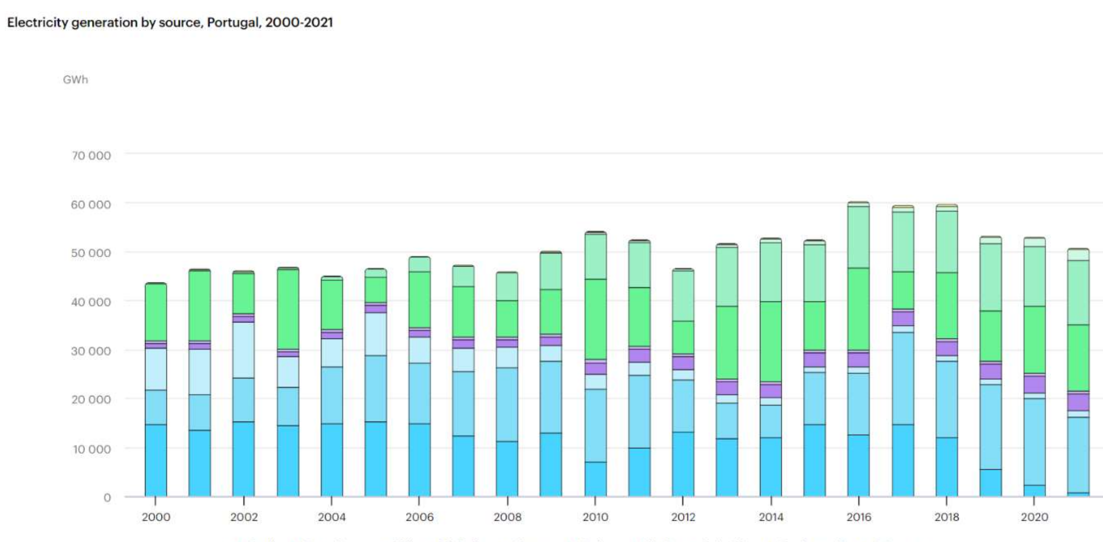


Figure 2. Electricity Generation by Source, Portugal, 2000-2021. Source: IEA.

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).

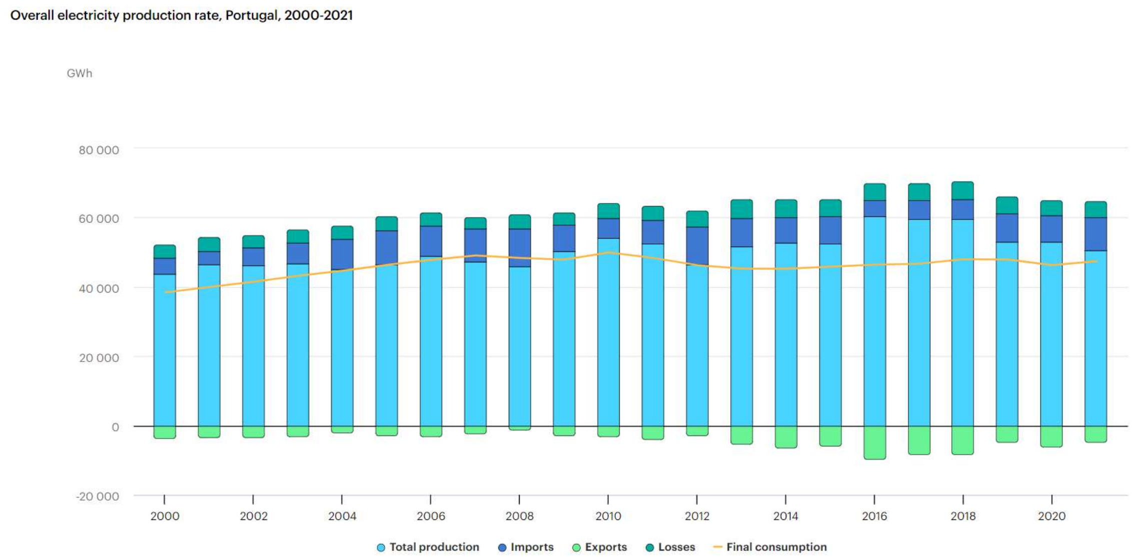


Figure 3. Overall Electricity Production Rate, Portugal, 2000-2021. Source: IEA.

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 31st to 9 AM on November 6th, 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 1st and 9 AM on November 5th, 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.

Furthermore, from 10 PM on October 31st to 9 AM on November 6th, 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.

- 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 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 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.

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 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 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_2 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.

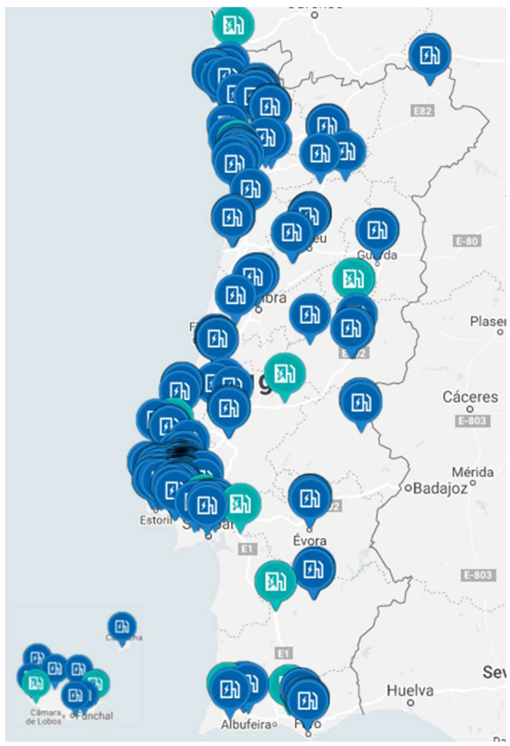


Figure 4. Distribution of CPEs in Portugal.

Electric load forecasting is pivotal in data-driven techniques for energy distribution optimization, ensuring efficient, reliable supply and supporting the intricate planning and operational demands of modern power systems. Accurate predictions of electricity demand are essential for allocating resources effectively and maintaining grid stability, particularly as grids become more complex and integrated with smart technologies.

Historically favored for their time series analysis, models like *ARIMA* have been instrumental in load forecasting by identifying and projecting historical load patterns into future trends. However, they often need to catch up in capturing the non-linear variability crucial for accommodating fluctuating renewable energy sources and evolving consumption patterns. The *SARIMAX* model addresses these limitations by incorporating exogenous factors, offering a more precise prediction of energy loads by considering influential variables like temperature, humidity, and consumer behavior.

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

6. Duarte Correia - Service Continuity Impacting Remote Services

6.1. Introduction

This study investigates how completed network connection requests impact service continuity indicators in the changing landscape of utility services, where proper management of network connection requests plays a critical role in guaranteeing continuous operation. These indicators are critical in maintaining the dependability and efficiency of utility services. With utility networks becoming more complex and facing increased connectivity needs, it is critical to understand the interplay between these demands and service delivery to maintain infrastructure resilience and customer satisfaction.

Rising remote work patterns and adopting renewable energy are examples of evolving utility service management challenges and opportunities. As a result, this dissertation explores how variations in network connection requests correspond with service continuity measures and remote work order trends, implying a broader shift toward data-driven ways to evaluate and enhance utility services.

Considering theoretical findings from the literature, this study develops two key assumptions. The first hypothesis suggests a direct relationship exists between the amount of network connection requests executed, changes in service continuity indicators, and the frequency of remote work orders. This hypothesis is based on the emphasis in the literature on the association between increased network activity and greater service robustness. The second hypothesis suggests that communities with a history of service continuity challenges will have more network connection requests, indicating a trend in which increased connectivity efforts serve as a response mechanism to earlier service issues.

This thesis uses *Regression analysis* and *Granger causality* testing to quantify and understand the complex relationships between network connection and service continuity indicators. The regression models are designed to evaluate the direct impact of network connection requests on service continuity and remote work orders. On the other hand, the *Granger* causality tests look at the temporal and potential causal relationships between these variables.

Based on data from the E-Redes Open Data Portal, the study discovers that, while there is a positive relationship between executed network connection requests and the volume of remote work orders, the influence on service continuity indicators is minor. In some instances, an increase in network connection requests correlates with improved service indicators, but in others, it suggests potential system stresses. These findings contribute critical insights to the utility service management field, enabling infrastructure construction and service optimization methods to enhance the reliability and efficiency of utility services for end users.

6.2 Literature Review

The rapid evolution of technology and operational challenges mark electric power distribution and service continuity in today's era. The literature in this domain offers a multidimensional view, encompassing technological innovation, operational resilience, cybersecurity, and the integration of big data analytics.

Santos, Pereira, and Machado Ferreira's (2014) exploration of distribution network optimization through computational tools underscores the technological advancements in network operations. They demonstrate how simulation tools can optimize network operations, aligning with the industry's trend toward computational solutions for operational excellence. However, their study opens avenues for further exploration into how network connection requests directly impact service continuity indicators.

Parise et al. (2016) delve into the operational efficiency within container terminal areas, drawing parallels with the resilience required in power distribution networks. Their dual-source electrical architecture suggests an understanding of the need for resilient infrastructure but needs to examine the effects of network connection requests on service continuity.

Amin (2012) addresses the growing cybersecurity concerns in smart grid environments. The study advocates for security as an integral component of smart grid architecture, highlighting the shift toward securing interconnected control systems. While this research is crucial in understanding secure control systems, it needs a more direct focus on the impacts of these cybersecurity measures on the dynamics between network connection requests and service continuity indicators.

Qiao et al. (2023) highlight the increasing significance of big data technologies in electric distribution data analysis. Their approach to modern storage systems and frameworks like Apache Spark mirrors the sector's adaptation to big data analytics. This shift hints at a potential research gap examining how executed network connection requests influence service continuity indicators within a data-rich environment.

Combining these studies presents a holistic strategy for electric power distribution, emphasizing operational improvements, infrastructure resilience, cybersecurity, and big data analytics. However, there needs to be more research that closely examines the relationship between executed network connection requests and service continuity indicators, especially considering the growing reliance on renewable energy and smart grid technologies.

This leads to the focused research area, which aims to tie this gap by analyzing how executed network connection requests influence service continuity indicators and remote work order trends. The next section of this research delves into the empirical evaluation of these relationships, contributing to a more comprehensive understanding of the operational dynamics within the utility sector. This exploration is particularly relevant in Portugal's evolving energy landscape, where integrating renewable energy sources and technological advancements is pivotal.

6.3 Research Questions & Hypothesis Development

Drawing insights from the reviewed literature, this study aims to explore an under-researched area by posing a critical research question:

Research Question: "*How do executed network connection requests influence service continuity indicators and current trends in remote work orders?*"

Hypotheses:

1. The volume of executed network connection requests directly impacts service continuity indicators and the number of remote work orders.

1a. Higher executed network connection requests correspond to improved service continuity indicators and an increase in remote work orders.

1b. Municipalities with historically poorer service continuity indicators have a higher volume of executed network connection requests.

6.4 Methods

6.4.1 Research Design

This study uses a quantitative research strategy, utilizing different statistical methodologies to understand better the relationship between executed network connection requests, service continuity indicators, and remote work order volume. Its objective is to quantify and test the interactions and effects of these variables across time to provide a thorough knowledge of their interdependence within the utility sector.

The research is built around regression models chosen for their ability to capture and quantify relationships between dependent and independent variables. *Linear Regression* allows for an early assessment of the potential linear relationships in the data.

The simplicity of *Linear Regression* is advantageous for the precise interpretation of direct relationships; however, it may not sufficiently capture more complicated, non-linear interactions. *Polynomial Regression* was also used to address potential non-linearities in the data. This model enhances *Linear Regression*'s capabilities by integrating polynomial terms, allowing the investigation of curved correlations between variables.

Because of its capacity to capture non-linear interactions and avoid overfitting, the *Random Forest* regression model emerged as the most appropriate model for this diverse dataset. It processed a great deal of data efficiently, outperforming simpler models in predicted accuracy and offering a better insight into the operational dynamics of utility services.

The *Random Forest* model outperformed the others, as evidenced by greater R-squared values and lower error metrics, establishing it as the best fit for this research. This methodological basis led to an in-depth look at how network connection demands affect service delivery and operational efficiency in the energy sector. In conclusion, the quantitative method, which the robust *Random Forest* regression model supports, provides a solid platform for examining the complicated interactions within the operational data of the utility sector. It enables a better understanding of the elements that influence service continuity and the effects of network connection requests on operational efficiency and service delivery.

6.4.2 Data Selection

The research used for this thesis was based on a carefully selected collection of secondary data gathered from E-Redes, Portugal's primary energy DSO. This information helped decipher the dynamics of executed network connection requests, service continuity indicators, and emerging patterns in remote work orders. The datasets were chosen for their extensive coverage of the utility sector's operational aspects and their fit with the research goals and hypotheses.

The *Service Continuity Indicators* dataset served as the foundation of the analysis. This dataset is a comprehensive collection of measures representing electrical services operational performance and dependability in various municipalities. It includes standardized reliability indicators such as the System *Average Interruption Frequency Index (SAIFI)* and the *System Average Interruption Duration Index (SAIDI),* which provide a quantitative measure of the frequency and duration of power outages experienced by consumers across multiple voltage levels (high, medium, and low).

In addition to these, the dataset includes the *Momentary Average Interruption Frequency Index (MAIFI)*, that indicates power supply stability, and the *Total Energy Index of Power Interruptions (TIEPI)* for medium tension, which provides information on the total duration of power outages. The Energy Not Supplied (END) statistic, denoted in megawatt-hours, quantifies the volume of electricity that could not be delivered owing to service interruptions. This is crucial in determining the true impact of outages on end-users.

The second dataset, *Remote Work Orders* was used to investigate the operation of utility services. It supplied a detailed monthly record of remote service orders broken down by district, municipality, and parish, revealing service demand and fulfillment patterns across areas. This dataset enabled the investigation of trends and seasonal variations in remote service demand, which is becoming increasingly important in developing grid demands and shifting toward remote service capabilities.

Finally, the *Executed Network Connection Requests* dataset provided valuable insights into network connectivity demand by showing the number of connections fulfilled biannually. This dataset acted as a barometer for operational workload and network expansion in the utility sector, highlighting significant patterns over time.

In their entirety, these datasets provided an abundance of information for solid research. They were chosen not only for their relevance and fit with the research aims but also for the data's validity and dependability, given its provenance from a credible and reliable source. While the datasets provide a thorough overview of the utility sector's operational indicators, their secondary nature means that the research is necessarily retrospective and dependent on the accuracy and completeness of the collected data.

6.5 Data Analysis

6.5.1 Descriptive Analysis

This section identifies patterns, correlations, and outliers that are not obvious from tabular data alone by using various representations.

The histograms of service continuity indicators show how various metrics are distributed across the dataset. The histograms in Appendix A for the *saifi at num, said at min, and mafia at num* indicators, for example, indicate a high frequency of low values, indicating that power outages are quite infrequent. On the other hand, the histogram for *end mt MWh* shows a broader variation of values, showing variability in the energy not supplied among municipalities or time periods. This spread may indicate places or periods where service disruption is more severe, necessitating further study.

Appendix I shows scatter plots that compare each service continuity indicator to executed network connection requests and show the correlations between these variables. The scatter plot for *Saifi at num* vs. *executed network connection requests*, for example, may provide a slight positive trend, implying that as network connections increase, so may the *SAIFI* at a high-tension level, which is in line with the first hypothesis. In contrast, the dense clustering at lower levels suggests that increasing connection requests have little effect on the frequency of disruptions in many localities.

The bar charts that rank municipalities based on the volume of remote work orders, worst service continuity indicators, and most executed network requests (in Appendix B, C, D and E) indicate places with the most significant operational demand and severe service continuity difficulties. The top two municipalities with the worst service score are Manteigas and Torre de Moncovo, while the top two in the number of executed network requests are Guimarães and Braga. This does not prove the hypothesis that municipalities with worse service have more executed network requests. However, this comparison must be fairer as Guimarães and Braga would have more requests due to increased population density and infrastructure demands. This does not mean poorer service continuity has no impact on connection requests, a topic that will be further explored in this thesis.

Yearly trend graphs for *end mt mwh* indicators in Appendix F and composite trends for *num* and *min* indicators in Appendix G and H provide insight into how these service continuity parameters have evolved over time. A strong peak or dip in these trends may indicate a specific year with significant events influencing service quality, such as infrastructure investments, natural disasters, or changes in utility management procedures.

6.5.2 Correlation Analysis

To better understand the relationship between *executed network connection requests*, *service continuity* indicators, and *remote work orders*, two statistical methods were employed:

The *Pearson Correlation* coefficient, known for its ability to detect linear relationships between continuous variables, was calculated to determine the strength of the linear association between executed network connection requests and service continuity metrics such as *SAIFI, SAIDI,* and *END.* The *Correlation Coefficient's p-value* assisted us in determining whether the observed associations were statistically significant or if they may be attributed to random variation.

A correlation matrix was also created to display these links simultaneously across several service continuity measures. This matrix validated the individual *Pearson Correlation* findings and provided a larger perspective on how these metrics are related to one another, boosting our knowledge of the dataset's internal structure.

Spearman's Rank Correlation was used as a supplemental non-parametric measure due to the potential limitations of *Pearson's Correlation*, such as including outliers or a non-normal data distribution. This method evaluates monotonic relationships based on data point rank order, which is especially useful for evaluating non-linear or non-parametric data. *Spearman's* analysis verified the *Pearson* results in this investigation, providing robustness against outliers and validating the direction and strength of the *Pearson*-identified associations.

The combined use of both correlation methodologies, aided by the visual aid of the correlation matrix, enabled a sophisticated interpretation of our data. The *Pearson* method sheds light on the linear relationships. At the same time, *Spearman's* methodology verified that our conclusions were robust to data distribution, resulting in a well-rounded understanding of the underlying patterns in our datasets.

6.5.3 Regression Analysis

Regression analysis was used in this study to investigate the relationship between *executed network connection requests* (independent variables) and *service continuity indicators* (dependent variables) across municipalities.

Ordinary Least Squares Regression (OLS): OLS regression was used to evaluate the relationship between the volume of executed network connection requests and various service continuity indicators such as *SAIFI, SAIDI,* and *END*.

The analysis enabled the forecasting of network connectivity's impact on service continuity, effectively controlling for the influence of other variables present in the datasets. Including a constant in the (OLS) model ensured the recording of the baseline level of service continuity when network connections were absent, providing a more accurate estimation. The model summary offered critical diagnostics, revealing the model's explanatory power and the importance of each predictor.

***Linear Regression* Model:** The *Linear Regression* model was used to understand better the impact of network connections on the number of remote work orders. This method is beneficial for simulating the relationship between a dependent variable (the number of *remote work orders*) and one or more independent variables (*executed network connection requests*). MSE and the coefficient of determination (*R-squared*) were used to assess the model's performance. By evaluating the average squared difference between actual and predicted values, the MSE measured the model's accuracy. R-squared calculated how much variation in remote work orders could be explained by network connection requests, indicating the model's overall efficacy.

***Polynomial Regression* Model:** A polynomial regression model was used to capture potential non-linear patterns that the linear model may have missed. This model is especially effective for discovering possible curvilinear correlations between the amount of performed *network connection requests* and *service continuity indicators*. The *polynomial regression* coefficients provided insight into the data's complex, non-linear interactions, enabling a more nuanced understanding of how network connection requests might affect service continuity indicators under varied scenarios.

***Lasso Regression* Model:** Another method used was the *Lasso (Least Absolute Shrinkage and Selection Operator)* regression. This model is useful for datasets with many features since it uses regularization to avoid overfitting by punishing the absolute magnitude of the regression coefficients. In our situation, it aided in determining which features are most relevant for predicting service continuity indicators and remote work orders, effectively decreasing the coefficients of the less critical variables to zero and thus conducting feature selection.

***Random Forest Regression* Model:** The *Random Forest* regression model, a technique for ensemble learning noted for its excellent accuracy and resilience, was also included in our investigation. Unlike simpler models, *Random Forest* builds many Decision Trees and integrates them to get a more accurate and reliable prediction. It was beneficial for our complicated dataset since it could manage the enormous number of input variables without requiring feature elimination, as evidenced by its superior performance compared to other models. The *Random Forest* model offered the best-predicted accuracy and highlighted the most influential elements, such as *end mt mwh*, which had the highest feature importance score, demonstrating its critical role in predicting service continuity.

6.5.4 Results

The application of *Pearson and Spearman correlation* analyses began to substantiate the primary hypothesis that the volume of executed network connection requests significantly influences both service continuity indicators and the volume of remote work orders. The *Pearson Correlation Analysis*, yielding a coefficient of 0.5436 for service continuity indicators and 0.1826 for remote work orders, with corresponding p-values of 0.0, indicates a positive linear relationship between network connection requests and both variables. This suggests that as network connection requests increase, there is a noticeable uptick in service continuity metrics and the number of remote work orders.

Similarly, *Spearman's Rank Correlation* results, with coefficients of 0.3955 for service continuity indicators and 0.2021 for remote work orders, reinforce this finding. Considering the rank-ordered data, the *Spearman* analysis supports the notion of a positive association, albeit indicating a moderate relationship for service continuity and a weaker one for remote work orders.

These results support the hypothesis, showing a statistically significant relationship between the volume of executed network connection requests and critical service indicators. However, the varying strengths of these correlations also highlight the complexity of these relationships, suggesting that other factors may also play significant roles in influencing these service metrics.

Table 1. OLS Regression Results – Energy Not Delivered.

OLS Regression Results	
Dependent Variable	Energy Not Delivered "end mt mwh"
Method	Least Squared
Feature	**Coefficient**
R-squared	0.296
Adjusted R-squared	0.295

The regression analysis presents significant findings that align with the proposed hypotheses. The OLS regression results for service continuity indicators, represented by *end mt mwh* reveal an R-squared of 0.296, which indicates that approximately 29.6% of the variability in service continuity can be explained by executed network connection requests. The model outlines a positive coefficient (0.0300) for the network connection requests, which suggests that an increase in these requests is associated with a notable increase in the energy not supplied, pointing to a potential area for improving service continuity measures. This finding does support the hypothesis that the volume of executed network connection impacts the service continuity indicators, however, it does not support the sub-hypothesis that more requests would correspond to improved service continuity indicators.

Table 2. Statistical Results - Energy Not Delivered.

	Const	Executed Network Requests
Coefficient	0.4220	0.0300
Standard Deviation Error	0.219	0.001
t	1.927	43.186
P>|t|	0.054	0.000
[0.025	-0.007	0.029
0.975]	0.851	0.031

The regression equation from this analysis is:

*End MT MWh = 0.4220 + (0.0300 * Executed Network Connection Requests)*

Table 3. OLS Regression Results - Number of Orders.

OLS Regression Results	
Dependent Variable	Number of Orders
Method	Least Squared
Feature	**Coefficient**
R-squared	0.033
Adjusted R-squared	0.033

For remote work orders, the *OLS regression* results display a lower R-squared value of 0.033, suggesting that executed network connection requests have a more minor but still significant influence on the number of remote work orders. The positive coefficient (0.0194) indicates that as the number of executed network connection requests rises, so does the number of remote work orders, albeit to a lesser extent than service continuity indicators, which supports the second half of our sub-hypothesis.

Table 4. Statistical Results - Number of Orders.

	Const	**Executed Network Requests**
Coefficient	10.1352	0.0194
Standard Deviation Error	0.142	0.000
t	71.155	75.267
P>|t|	0.000	0.000
[0.025	9.856	0.019
0.975]	10.414	0.020

The regression equation for remote work orders is:

*Number of Orders = 10.1352 + (0.0194 * Executed Network Connection Requests)*

These findings partially support the hypothesis that increased network connection requests correlate with service continuity indicators and remote work order volume changes. While there is a positive relationship, the strength of this relationship varies, with a more pronounced impact observed on service continuity indicators than on remote work orders. This nuanced outcome suggests that while network connection requests are indeed a factor in service continuity and demand for remote services, they are not the sole determinant, and other variables may also play significant roles.

Linear Regression was the first phase, providing a foundational understanding of the potential linear relationships within the datasets. While its clarity was provided by its simplicity, the results revealed a very moderate explanatory power for both service continuity indicators (R^2 = 0.0156) and remote work orders (R^2 = 0.0332), with MSE of 1475.46 and 1655.02, respectively. This revealed that the link between network connection requests and specific operational measures was not linear and could be impacted by other latent factors.

Polynomial Regression, with its ability to capture non-linearities, presented a more intricate approach than linear models. However, this sophistication translated into a somewhat obscure understanding, especially for service continuity indicators. The polynomial model yielded an R^2 of 0.1920 but a higher MSE of 48373.16. Its equation: y = 124.08338102206244 + 13.48048738x - 0.04011091x^2 indicates the nuanced relationship between *end mt mwh* and *Executed Network Connection Requests*. Similarly, for remote work orders, the model showed a slight increase in R^2 to 0.0557 but an elevated MSE of 1616.49, with its equation being $y = 2.950389899606863 + 0.0615893821x - 0.0000286059928x^2$. These results underline the complexity of predicting service continuity and remote work trends using *Polynomial Regression.*

Polynomial Regression was introduced to handle non-linearities that its linear equivalent might have ignored. Despite its more advanced approach, the polynomial model did not considerably improve comprehension, notably for service continuity indicators, with an even larger MSE of 48373.16 and a small R^2 of 0.1920. For remote work orders, the polynomial model provided a minor increase in R^2 to 0.0557 but also an elevated MSE of 1616.49. The plots in Appendix J and K show how the polynomial models fared.

Regularization was introduced in *Lasso Regression* to improve model performance by penalizing the coefficients' absolute magnitude, hence reducing overfitting. However, for both service continuity indicators (R^2 = 0.1542, MSE = 50642.28) and remote work orders (R^2 = 0.0332, MSE = 1655.02), the Lasso model performed similarly to the linear model, demonstrating that simply penalizing coefficients was insufficient to reflect the complexity of the data.

The *Random Forest* model stood out as the best performer of all the models used, as seen in Appendix L and M. Even though its prediction capability was not exceptional, as indicated by the plots, it had an ensemble technique that helped it capture complex, non-linear relationships between variables without requiring explicit feature transformation. The *Random Forest* model produced an R^2 of 0.5380, significantly higher than the other models, and a lower MSE of 27662.41, indicating a much stronger fit to the data. The model outperformed the others for remote work orders with an R^2 of 0.2416 and an MSE of 1298.26.

The *Random Forest* model's feature importance gave important insights, identifying *end mt mwh* as the most significant predictor as seen in Appendix N. This demonstrates *Random Forest*'s robustness in managing feature selection and the significance of *end mt mwh* as a critical factor in predicting service continuity.

The *Granger* causality test was used to investigate the relationship between the number of executed network connection requests and the amount of END, as assessed by the *end mt mwh* indicator. This statistical method seeks to determine whether one time series can foretell another, which aids in understanding the directional influence of network connections on service continuity.

The *Granger* causality test findings showed no significant predictive link between the executed network connection requests and *end mt mwh* across multiple lag times. There is insufficient statistical evidence to indicate a *Granger*-causal link in either direction, with p-values over the 0.05 level for all lag intervals studied. h Here is a breakdown of the *Granger* causality test results by the number of lags:

Lag 1: The p-value of 0.1951 indicates no *Granger* causality between 'end mt mwh' and executed network connection requests.

Lag 2: Similarly, a p-value of 0.3541 at lag 2 reinforces the absence of *Granger* causality.

Lag 3: At lag 3, the p-value of 0.2239 indicates no *Granger*-causal relationship.

Lag 4: Finally, a p-value of 0.2712 at lag 4 confirms the lack of *Granger* causality.

These *Granger* causality test results indicate that service continuity indications such as *end mt mwh* do not *Granger*-cause changes in executed network connection requests within the scope of the data and the selected lag intervals. This means that increases in network connection requests do not statistically anticipate energy changes not delivered owing to service outages, at least not in a way that this test can detect. As a result, these findings do not support the sub-hypothesis that municipalities with historically poor service continuity indicators (higher *end mt mwh*) would have a greater volume of executed network connection requests due to *Granger* causality.

6.6 Discussion

This study's analysis revealed intricate dynamics between executed network connection requests and their impact on service continuity indicators and remote work orders. The findings imply that increased quantities of network connection requests are not always related to increases in service continuity indicators. While most indicators improved, others, such as *System Average Interruption Duration Index* and *END*, deteriorated, showing a more complex interplay of components than previously predicted. This intricacy calls into question existing models that expect a direct relationship between more infrastructure investment and improved service outcomes.

The positive correlation between network connection requests and *end mt mwh* is especially telling, implying that increased connection requests may be associated with more energy not being given. This could indicate a strain on existing infrastructure or difficulties in managing the increased load, which have severe implications for practice. It indicates that utility companies should assess not just the number but also the quality of infrastructure expenditures to ensure that network development does not jeopardize service continuity.

The data for remote work orders support the premise of a beneficial impact. According to the findings, as the number of network connection requests grows, so does the volume of remote work orders, demonstrating a clear relationship between network expansion or activity and demand for associated services. This has practical consequences for utilities, emphasizing the significance of strategic planning in resource allocation, labor management, and remote service scheduling to meet escalating demand.

6.7 Limitations of the Study

The study's weaknesses are mostly connected to the data and its breadth. While the E-Redes statistics are vast, they are limited to a specific geographic and operational context—Portugal's electrical grid. As a result, without further examination, the findings may not be generalizable to other countries or energy sectors.

Furthermore, because the data is secondary, it is subject to the restrictions of how it was gathered, including any inherent biases or inaccuracies in the original data recording. Because the study relies on historical data, the quickly changing character of the utility industry, as driven by technological improvements and regulatory shifts, may need to be fully captured in the analysis.

6.8 Conclusion

The investigation in this thesis began to determine the impact of completed network connection requests on service continuity indicators and the changing patterns in remote work orders. Towards the conclusion of this analysis, it becomes clear that the relationship between infrastructure growth and service continuity is intricate and deviates from the linear patterns typically predicted by classical models.

These findings lead to a better knowledge of the operational difficulties and opportunities in the utilities sector. Increases in network connection requests have been shown to potentially cause service continuity challenges, as evidenced by the spike in *end mt mwh*. This realization necessitates a rethinking of current practices and emphasizes the need for utility companies to spend not only to expand their networks but also to improve the resilience and efficiency of these systems.

On the other hand, the positive association between executed network connection requests and the number of remote work orders indicates a strong demand for remote service capabilities, indicating a service sector responsive to infrastructure changes. This responsiveness is an encouraging sign that the industry is adapting to new operational paradigms and customer expectations.

As concluded, the initial theories only partially describe the complicated processes at work. While some expected links were observed, others were refuted or proved more complex than anticipated. This complexity adds to the comprehension and encourages additional investigation.

6.8.1 Future Research Directions

In the future, this research lays the framework for various lines of investigation:

Broader Geographic **Analysis:** Broadening the scope of the analysis to include data from other countries or areas may provide comparative insights and a more generalizable knowledge of the observed dynamics.

Incorporating real-time monitoring data could provide a more nuanced view of how service continuity indicators fluctuate with network connection requests.

Qualitative Research: By engaging with stakeholders in the utility industry, qualitative research could delve into the underlying causes of the observed patterns.

Examining the role of developing technologies such as smart grids and renewable energy sources shows how innovations affect the operating landscape.

Policy and Regulatory Implications: Understanding the effects of policy and regulatory changes on service continuity could help inform more effective sector governance.

Finally, this research should not be the end but rather a steppingstone in an ongoing discussion about the future of utility services. The findings could provide a framework for practitioners and theorists to build on as they traverse the complexity and obligations of maintaining dependable, efficient, and responsive utility services in an ever-changing world.

References

Aguiar, Carla. 2022. "Portugueses rendem-se à energia solar em plena crise energética." *Diário De Notícias*, August 8, 2022. https://www.dn.pt/dinheiro/portugueses-rendem-se-a-energia-solar-em-plena-crise-energetica-15078038.html.

Ahmad, Tanveer, Rafał Madoński, Dongdong Zhang, Chao Huang, and Asad Mujeeb. "Data-Driven Probabilistic Machine Learning in Sustainable Smart Energy/Smart Energy Systems: Key Developments, Challenges, and Future Research Opportunities in the Context of Smart Grid Paradigm." *Renewable & Sustainable Energy Reviews* 160 (May 1, 2022): 112128. https://doi.org/10.1016/j.rser.2022.112128.

Box George E. P Gwilym M Jenkins and Gregory C Reinsel. 1994. *Time Series Analysis: Forecasting and Control* Third ed. Englewood Cliffs N.J: Prentice Hall. https://search.worldcat.org/title/Time-series-analysis-:-forecasting-and-control/oclc/28888762.

Chandra, Ambarish, Sumeet Gulati, and Milind Kandlikar. 2010. "Green Drivers or Free Riders? An Analysis of Tax Rebates for Hybrid Vehicles." *Journal of Environmental Economics and Management* 60 (2): 78–93. https://doi.org/10.1016/j.jeem.2010.04.003.

Da Fonseca Monteiro Pereira, Rita Manuela. "Redes de Distribuição de Energia Elétrica de Média e Baixa Tensão : Estágio Na Helenos, S.A," June 27, 2018. https://comum.rcaap.pt/handle/10400.26/25475.

European Economic and Social Committee. 2023. "Energy – Producing and consuming own energy key to fighting climate change and energy poverty." November 22, 2023. https://www.eesc.europa.eu/pt/news-media/news/energy-producing-and-consuming-own-energy-key-fighting-climate-change-and-energy-poverty#related-content.

Goldenergy. 2022. “Incentivos e benefícios fiscais dos painéis solares.” *Goldenergy* (blog). March 22, 2022. https://goldenergy.pt/blog/autoconsumo/beneficios-fiscais-paineis-solares/.

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.

International Energy Agency. 2021. “Portugal 2021.” IEA. July 2021. https://www.iea.org/reports/portugal-2021.

International Energy Agency. 2022. “Portugal Electricity Security Policy.” IEA. June 30, 2022. https://www.iea.org/articles/portugal-electricity-security-policy.

Junfeng Qiao, Aihua Zhou, Lin Peng, Hua Gu, Xiaofeng Shen, Chenhong Huang, "Research and implementation of key technologies for electric distribution data analysis," 2023 Proc. SPIE 12703. https://spie.org/Publications/Proceedings/Paper/10.1117/12.2682795?SSO=1.

Mortimer, Benedict J., Christopher Hecht, Rafael Goldbeck, Dirk Uwe Sauer, and Rik W. De Doncker. 2022. “Electric Vehicle Public Charging Infrastructure Planning Using Real-World Charging Data.” *World Electric Vehicle Journal* 13 (6): 94. https://doi.org/10.3390/wevj13060094.

N. Bhusal, M. Abdelmalak, M. Kamruzzaman and M. Benidris, "Power System Resilience: Current Practices, Challenges, and Future Directions," in IEEE Access, vol. 8, pp. 18064-18086, 2020, https://ieeexplore.ieee.org/abstract/document/8966351.

Nogueira, Filipe Tomé Brites Ferreira. 2014. “Produção de Energia Por via Fotovoltaica Para Autoconsumo Instantâneo Com Ligação à Rede.” Universidade da Beira Interior.

Ribeiro, Paulo, José F. G. Mendes, and Rui a. R. Ramos. 2012. *Advances in Environment, Computational Chemistry and Bioscience: Electric Mobility in Portugal – Municipal Plans for Its Promotion*.

S. M. Amin, "Smart grid security, privacy, and resilient architectures: Opportunities and challenges," 2012 IEEE Power and Energy Society General Meeting, San Diego, CA, USA, 2012, pp. 1-2, https://ieeexplore.ieee.org/abstract/document/6345767.

Sierzchula, W., Bakker, S., Maat, K., & van Wee, B. (2014). The influence of financial incentives and charging infrastructure on EV adoption. https://www.sciencedirect.com/science/article/pii/S0301421514000822.

Sierzchula, William, Sjoerd Bakker, Kees Maat, and Bert Van Wee. 2014. "The Influence of Financial Incentives and Other Socio-Economic Factors on Electric Vehicle Adoption." *Energy Policy* 68 (May): 183–94. https://doi.org/10.1016/j.enpol.2014.01.043.

Suspiro, Ana. 2023. "Produção renovável abasteceu sozinha consumo nacional de eletricidade por 149 horas (cerca de seis dias)." *Observador*, November 6, 2023. https://observador.pt/2023/11/06/producao-renovavel-abasteceu-sozinha-consumo-nacional-de-eletricidade-por-149-horas-cerca-de-seis-dias/.

V. Santos, A. J. C. Pereira and C. M. Machado Ferreira, "Effective and simple measures to improve the Electric Power Distribution systems service continuity," 2014 49th International Universities Power Engineering Conference (UPEC), Cluj-Napoca, Romania,2014,pp.1-6, https://ieeexplore.ieee.org/abstract/document/6934767?casa_token=cfQZfE_hRewAAAAA:cjXQCBwCAbXNwweMKzVVG1sZHvcguYWN2oLLk4usVbAMFwX5b9cOfjMLqjKl9W8FM1Zs5i1i9fE.

Wee, Sherilyn, Makena Coffman, and Sumner La Croix. 2018. "Do Electric Vehicle Incentives Matter? Evidence from the 50 U.S. States." *Research Policy* 47 (9): 1601–10. https://doi.org/10.1016/j.respol.2018.05.003.

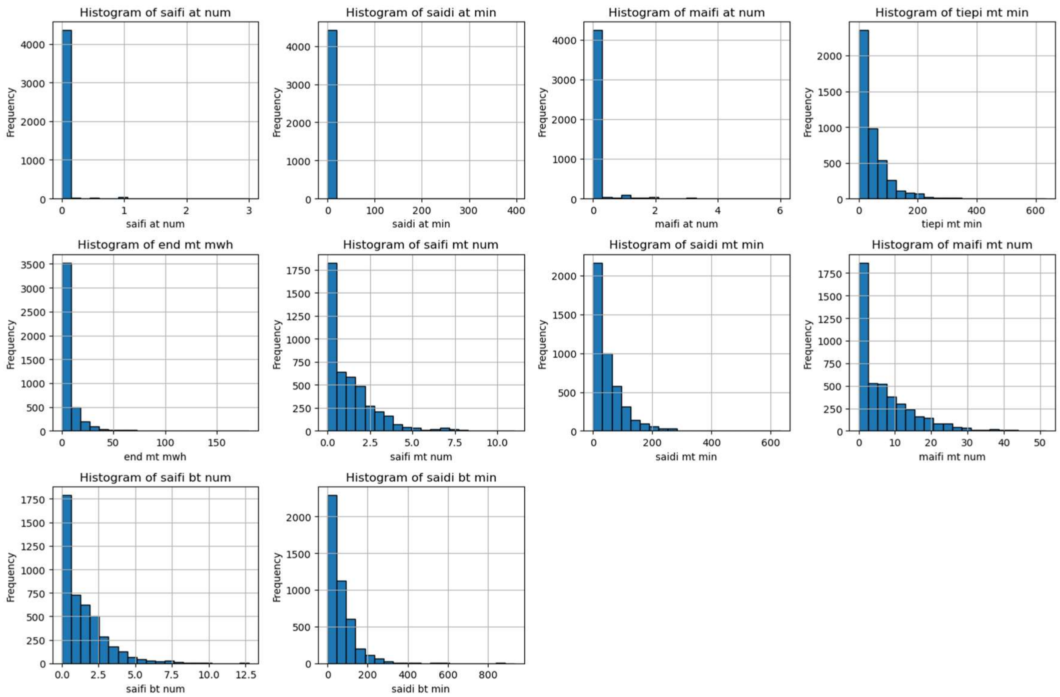
Zhang, Qi, Hailong Li, Lei Zhu, Pietro Elia Campana, Huihui Lu, Fredrik Wallin, and Qiang Sun. 2018. "Factors Influencing the Economics of Public Charging Infrastructures for EV – A Review." *Renewable & Sustainable Energy Review*

Zhou, Kaile, Chao Fu, and Shanlin Yang. "Big Data Driven Smart Energy Management: From Big Data to Big Insights." *Renewable & Sustainable Energy Reviews* 56 (April 1, 2016): 215–25. https://doi.org/10.1016/j.rser.2015.11.050.

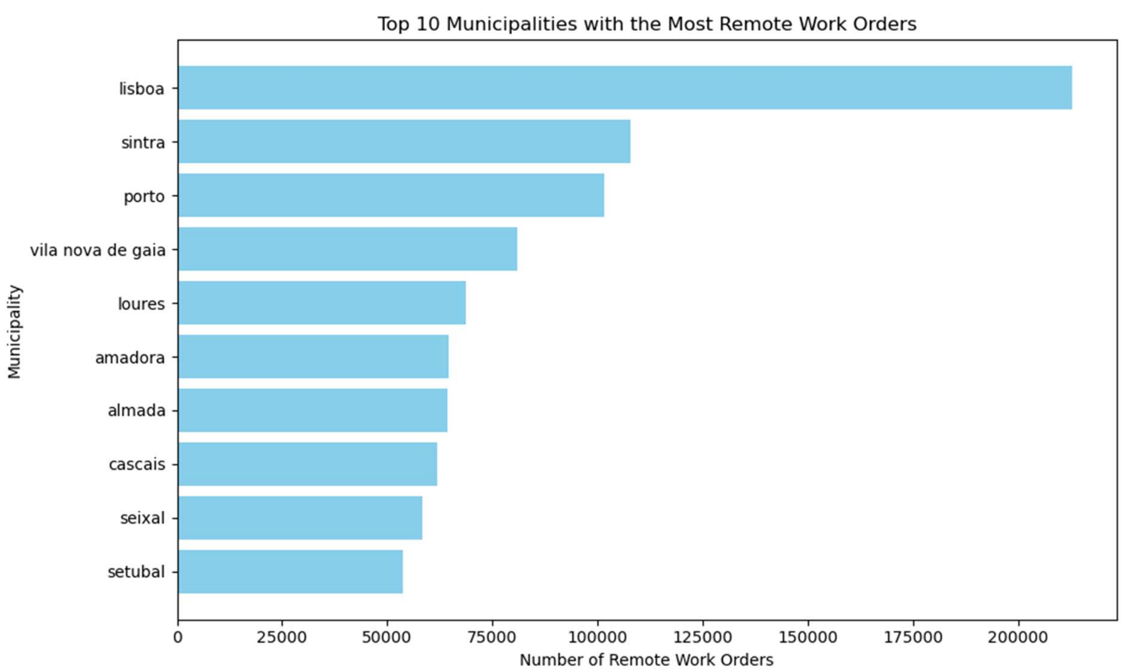
Zhou, Yan, Michael Wang, Han Hao, Linda Cooke Johnson, and Hewu Wang. 2014. "Plug-in Electric Vehicle Market Penetration and Incentives: A Global Review." *Mitigation and Adaptation Strategies for Global Change* 20 (5): 777–95. https://doi.org/10.1007/s11027-014-9611-2.

Appendix

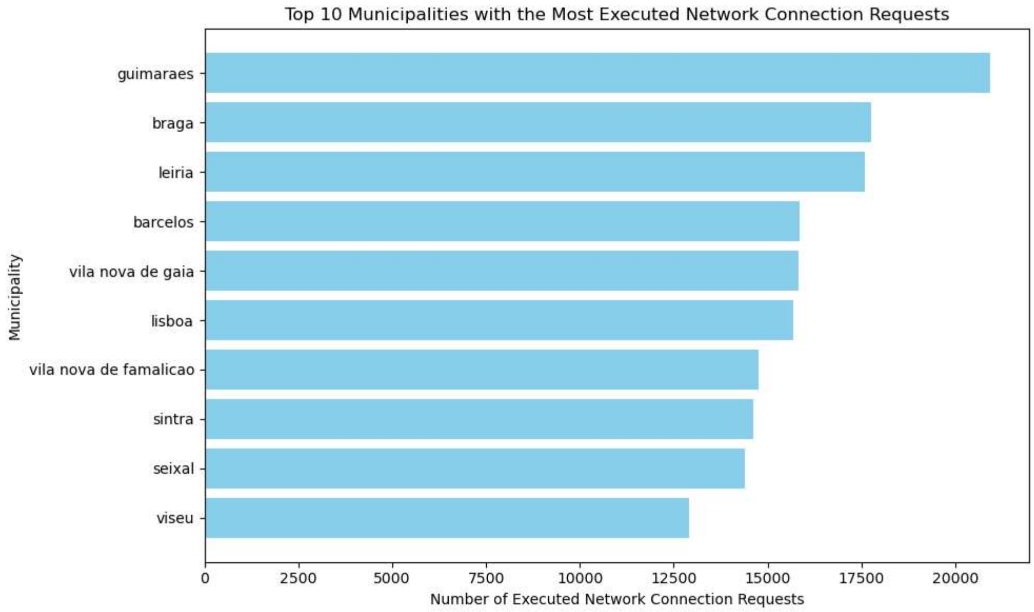
Appendix A. Frequency Histograms of Various Service Continuity Indicators.



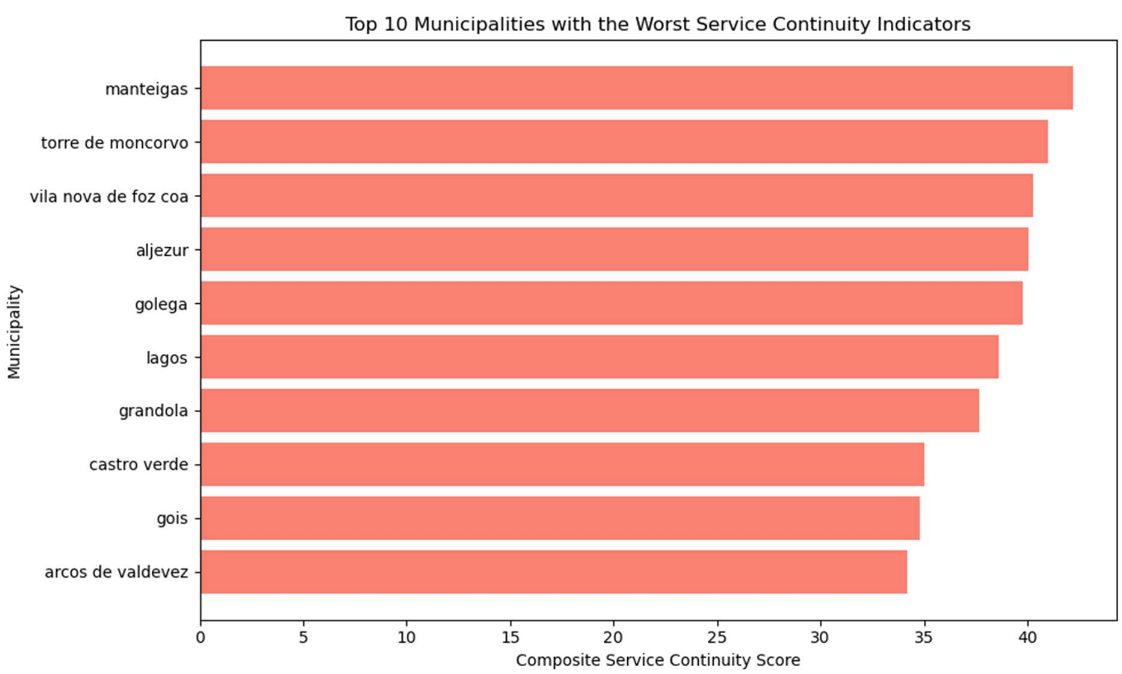
Appendix B. Top 10 Muncipalities with the Most Remote Work Orders.



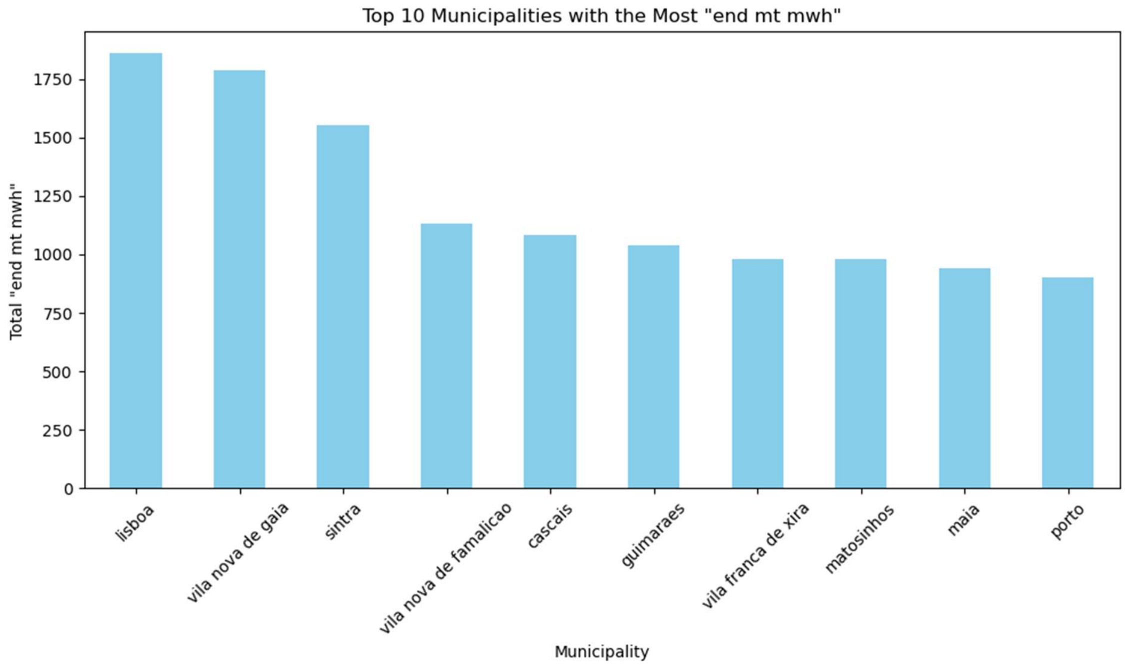
Appendix C. Top 10 Muncipalities with the Most Executed Network Requests.



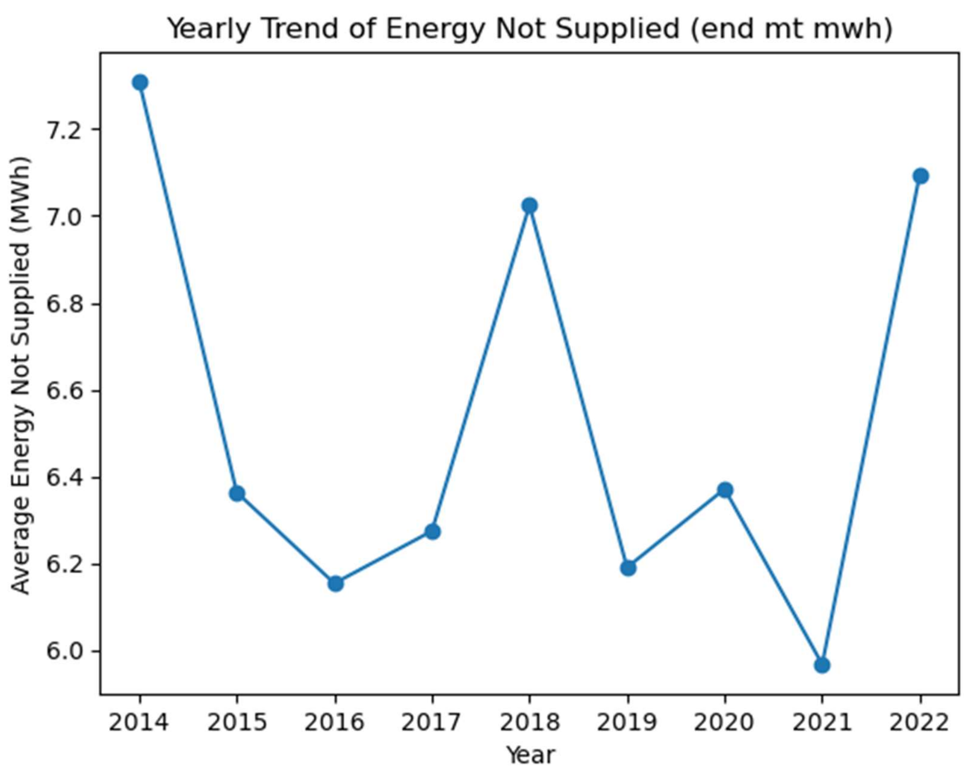
Appendix D. Top 10 Muncipalities with the Worst Service Score.



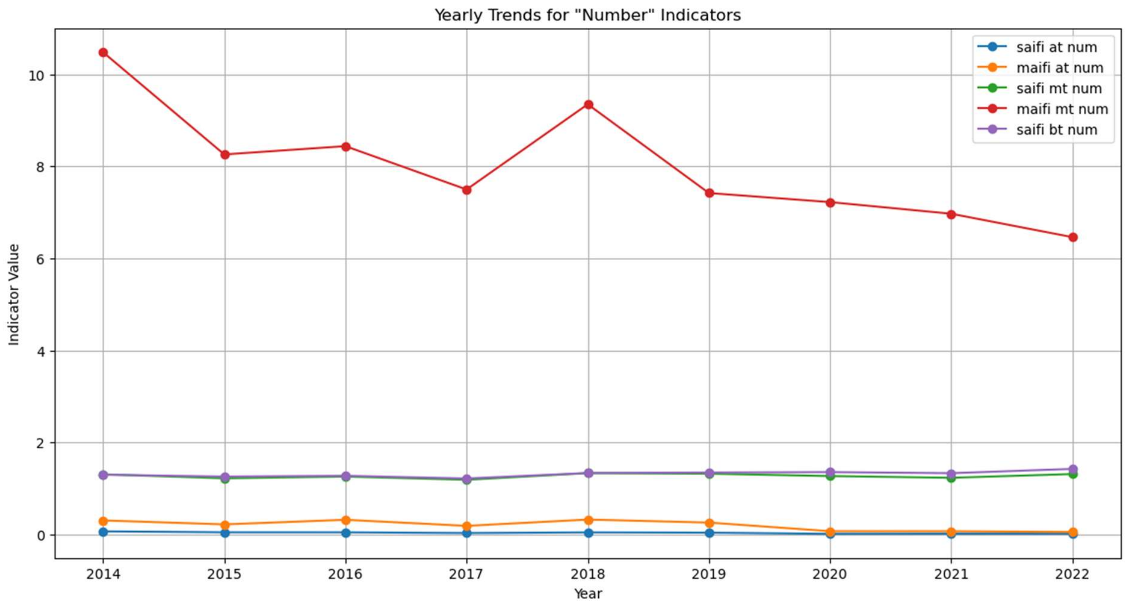
Appendix E. Top 10 Muncipalities with the Most Energy Not Delivered "end mt mwh".



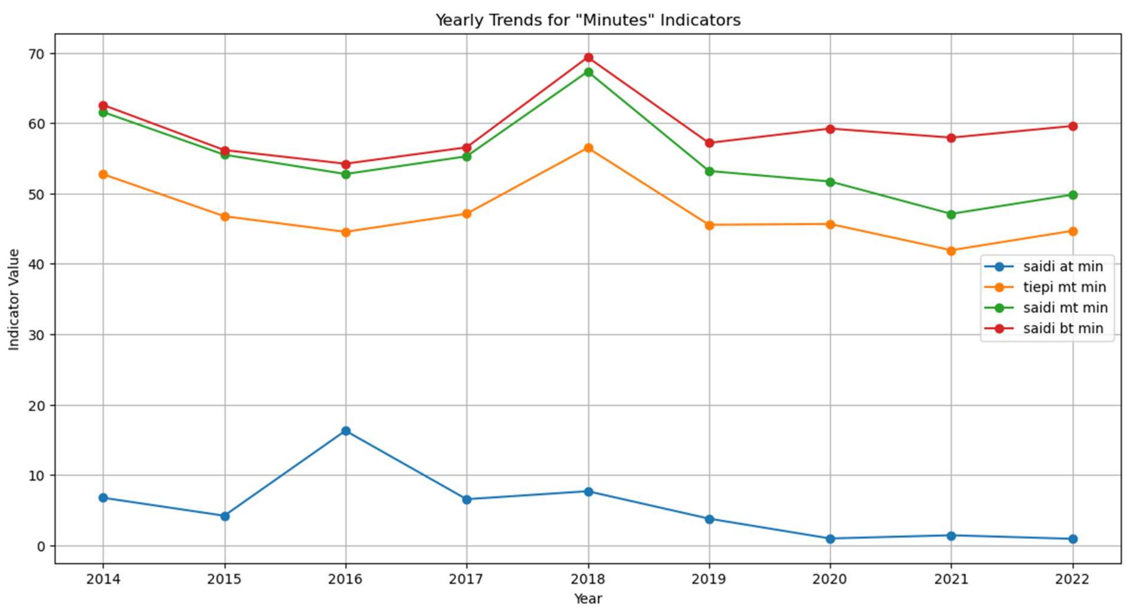
Appendix F. Yearly Trend for Energy Not Supplied.



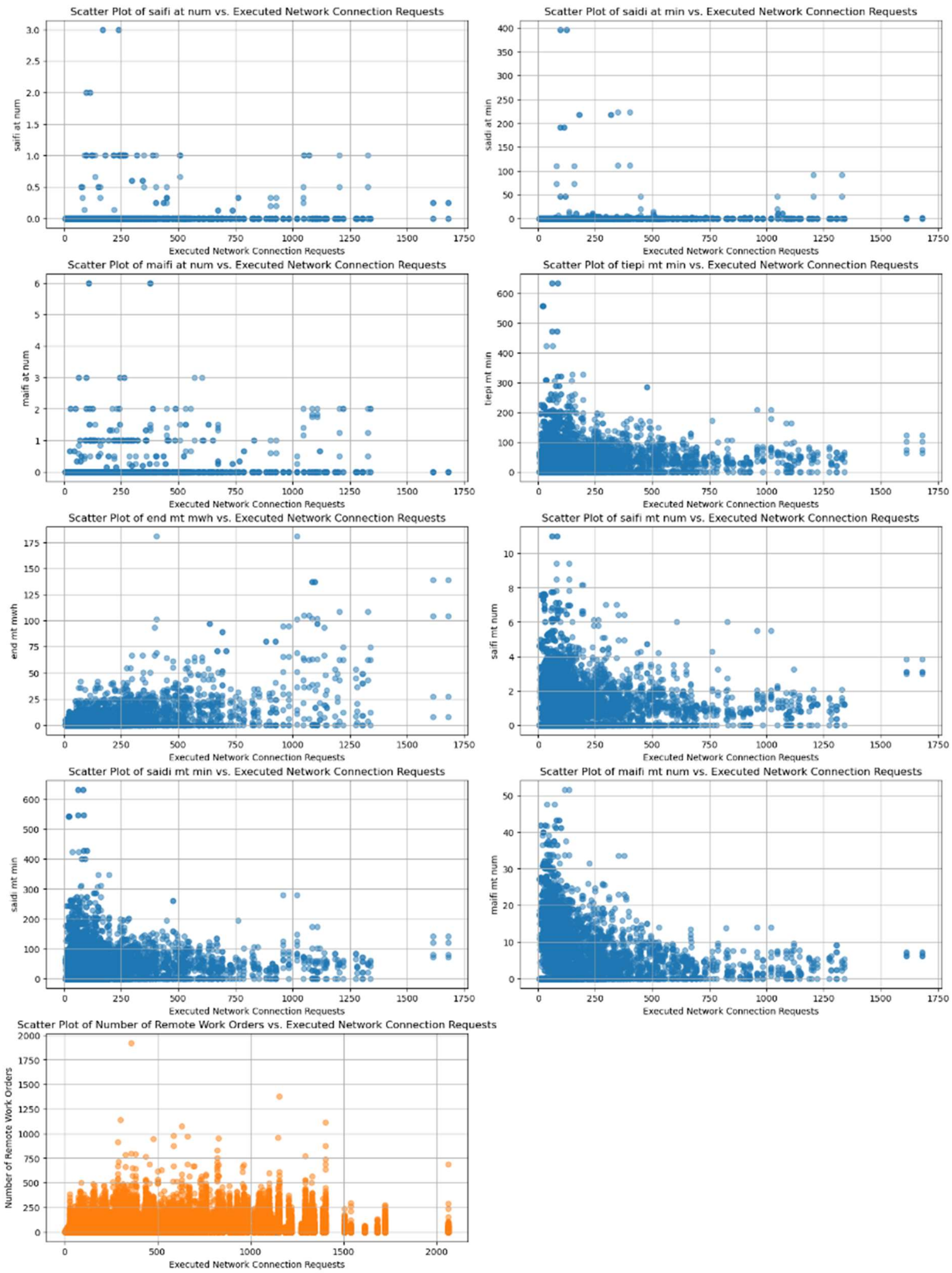
Appendix G. Yearly Trends for Service Indicators in Numbers.



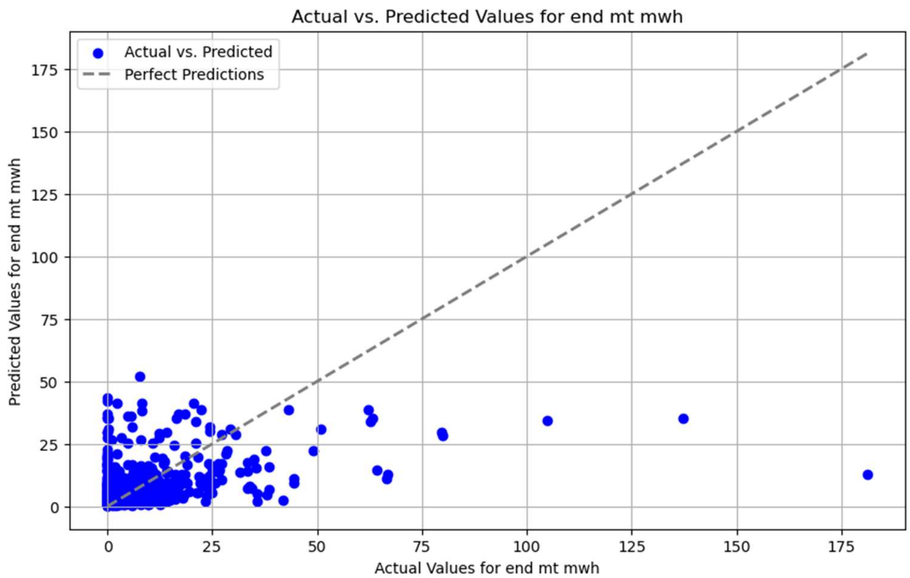
Appendix H. Yearly Trends for Service Indicators in minutes.



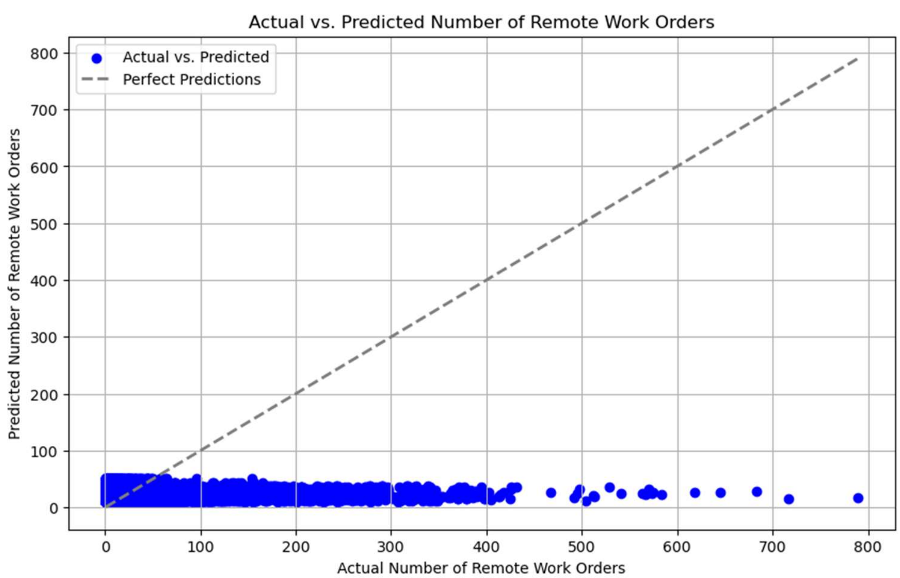
Appendix I. Scatter plots of Service Indicators vs Executed Network Requests.



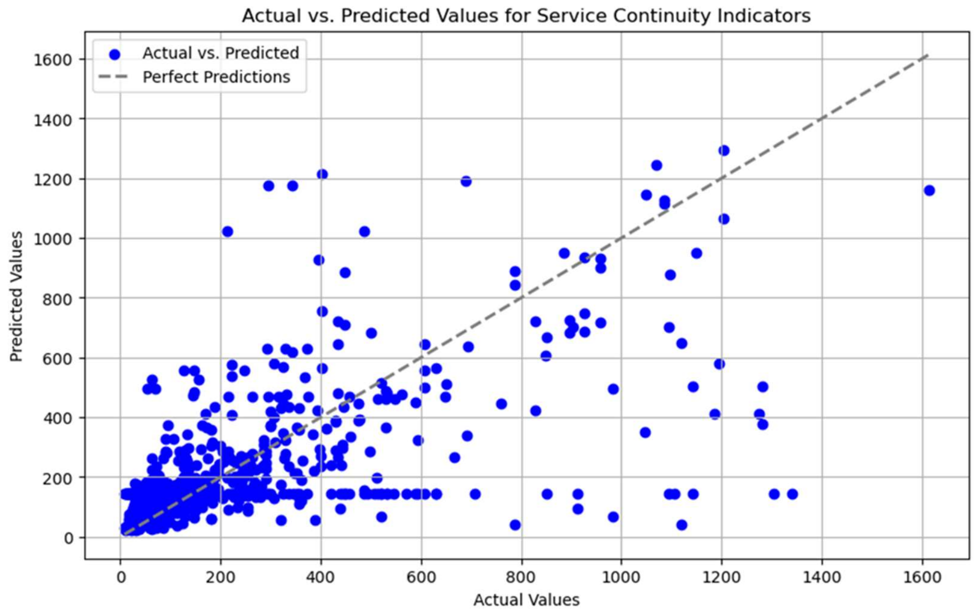
Appendix J. Comparison of Actual vs Predicted Values for Energy Not Delivered.



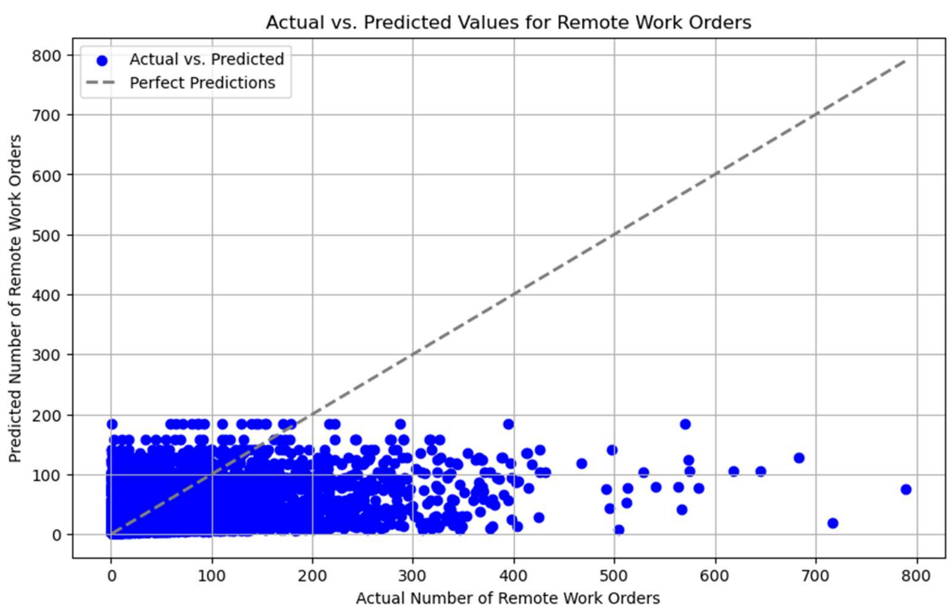
Appendix K. Comparison of Actual vs Predicted Number of Remote Work Orders.



Appendix L. Comparison of Actual vs Predicted Values for Energy Not Delivered.



Appendix M. Comparison of Actual vs Predicted Number of Remote Work Orders.



Appendix N. List of Feature Importance for Service Continuity Indicators on Random Forest Model.

```
Random Forest Feature Importance for Service Continuity Indicators (Ordered from Greatest to Lowes
t):
end mt mwh: 0.5314686437438018
maifi mt num: 0.15913333677211142
saidi bt min: 0.07509625759008087
saifi mt num: 0.061243374101129455
saifi bt num: 0.05329501982784629
tiepi mt min: 0.0512862583360167
saidi mt min: 0.04906720635035397
maifi at num: 0.016239858094488323
saifi at num: 0.0019868661439237925
saidi at min: 0.001183179040247526
```