Master's Thesis

MSc Energy for Smart Cities

Definition of a Digital Tool to Boost Photovoltaic Self-Consumption Projects in Energy Communities

MEMORY

Autor: Monika Kucharczyk

Director: Mónica Aragüés Peñalba

Company Supervisor: Gerard Laguna Benet

Convocation: September 2022

ETSEIB

Escola Tècnica Superior
d'Enginyeria Industrial de Barcelona

UPC

Abstract

Global climate change is one of the fundamental problems of our time. Its solution requires close, innovative, and successful cooperation between all governments. Policymakers have made significant attempts in recent years to reach consensus on globally binding climate objectives. The Paris Agreement, which was approved in December 2015, was the result of these studies. According to the report, stabilizing temperature increases at less than 2 degrees Celsius by 2050 is crucial to prevent the worst and most harmful impacts of a changing climate system.

The society is becoming more and more conscious of the need to alter our current energy system and switch from fossil fuels to renewable energy sources with the focus on energy efficiency, distributed generation and the concept of electrification. This energy revolution paves the way for new energy projects, in particular at the empowerment of smaller players in the energy market as well as an increase in decentralized renewable energy generation and consumption, so called prosumption. Furthermore, it opens up a possibility of the formation of local energy communities which are key to address the challenge of climate change.

This thesis focuses on developing a digital tool that will help with the boost of the solar photovoltaic self-consumption projects towards energy communities. Members of such initiatives, thanks to the current regulations, are now entitled to generate renewable energy without charges (including for their own consumption), store and sell their excess production with remuneration and perform peer-to-peer energy trading. The concept of an energy community brings number of benefits and opportunities for consumers which is described in detail in this thesis.

This work was developed as a part of European project called ePLANET which belongs to the Horizon 2020 initiative and was written under the supervision of BEE Group - an independent department of the Catalan research centre CIMNE. The main objective of this work is the collection of energy and gas consumption data for all the buildings of the project's pilot municipalities in Girona and matching it with the Spanish cadastre, with the use of the harmonization script developed in Python environment. Thanks to this process it is possible to characterize the buildings of different economic sectors and visualize the results on the GIS-based platform. Additionally, the obtained energy profiles of different buildings are the perfect foundation for the calculation of the size of energy community components and the optimization processes, as well as many other energy-related studies with a broader scope.

Contents

List of Figures

1 Glossary

A list of words and abbreviations used in this thesis may be found below.

- ACER: Agency for the Cooperation of Energy Regulators
- EC: Energy Communities
- PV: Photovoltaics
- kW: Kilowatt
- kWh: Kilowatt Hour
- MW: Megawatt
- MWh: Megawatt Hour
- GW: Gigawatt
- EU : European Union
- GIS: Geographical Information System
- PVGIS: Photovoltaic Geographical Information System
- PVSC: Photovoltaic Self Consumption
- SECAP: Sustainable Energy and Climate Action Plan
- TW: Terawatt
- DSO: Distribution System Operator
- TSO: Transmission System Operator
- NZEB: Nearly Zero Energy Buildings
- UN: United Nations
- SDGs: Sustainable Development Goals
- NECPs: National Energy and Climate Plans
- RD: Royal Dectreto
- H2020: Horizon 2020
- ETM: Energy Transition Measures

- CIMNE: Centre Internacional de Metodes Numerics en Enginyeria
- UNESCO: The United Nations Educational, Scientific and Cultural Organization
- LiDAR: Light Detection and Ranging
- DSM: Demand Side Management

2 Structure of the Thesis

The current Master's Thesis is structured in the following way:

- **Introduction**
An introduction to the general topic of a climate crisis and established energy goals. It also includes photovoltaic market analysis together with the current legislative framework. Another big part of this section is the aim and motivation of the thesis, with the description of the supervising company BEE Group, and the definition of the ePLANET project, which this work was part of.

- **Energy Communities**
It discusses the local energy association concept, as well as formal definitions, challenges, advantages, and disadvantages. This section also includes briefings to the Spanish regulations concerning self-consumption.

- **State of the Art**
Analysis of platforms and companies which have the similar goal of contributing to the creation of energy communities. It was divided into two different categories: GIS-based platforms and tools for the municipalities.

- **Particular Business Case**
The essence of this thesis. It provides an introduction to the studied case, as well as a detailed description of the methodology used for this project. This section outlines the operations performed on the dataframes: collection, preprocessing, and harmonization. The final part includes the results, its analysis, and discussion regarding the outcomes of the project. Budget analysis is also a fraction of this section.

- **Conclusions, Contributions and Future Work**
This section presents the main findings and contributions of this thesis with ideas for further development and future work.

3 Introduction

3.1 The Energy Goals

Due to human activity, the climate is warming up with dangerous negative effects. According to the United Nations (UN) climate change discussions, the global average temperature rise should not exceed 2 degrees, and should even be kept to a minimum of 1.5 degrees.

This goal was adopted by 196 Parties at **COP 21** in Paris, on 12 December 2015 and entered into force on 4 November 2016, and was called the **Paris Agreement** [1]. Countries want to reach the global greenhouse gas emissions peak as soon as feasible in order to build a climate-neutral world by the middle of the century in order to meet this long-term temperature objective. To keep within these restrictions, drastic and quick emission reductions are required. Adaptation is necessary since climate warming will persist despite attempts to reduce it.

In order to assist the transition from fossil fuels to cleaner energy and, more particularly, to fulfill the EU's Paris Agreement objectives for decreasing greenhouse gas emissions, the EU updated its energy policy framework in 2019. **The Clean energy for all Europeans** package deal, which is a new set of energy regulations, was a major step toward implementing the energy union strategy.

The package contains 8 new legislation that is based on suggestions from the Commission released in 2016. EU nations have 1-2 years to enact the new directives into national law following the political agreement reached by the EU Council and the European Parliament and the effective date of the various EU regulations. The environment, the economy, and consumers will all gain greatly from the new regulations. The Act further emphasizes EU leadership in combating global warming by coordinating these reforms at the EU level and significantly advances the EU's long-term aim to achieve carbon neutrality (net-zero emissions) by 2050 [2].

To achieve this level of carbon reduction the clean energy for all Europeans package deal is transversal across all the energy systems. For this, the package deal contemplates the energy performance in buildings, the increase of renewable energy, improving the energy efficiency, modifying the governance regulation and also redesigning of the electricity market.

- **Energy Performance in Buildings**
  Buildings are responsible for around 40% of energy consumption and 36% of CO_2 emissions in the EU, making them the single largest energy consumer in Europe. The EU can more easily accomplish its energy and climate goals by increasing the energy efficiency of buildings. The Energy Performance of Buildings Directive ((EU 2018/844)) updates and modifies numerous earlier laws (Directive 2010/31/EU) and defines specific measures for the building industry to address issues.

- **Renewable Energy**
The EU has set an aggressive, legally enforceable goal of 32% renewable energy sources in the EU's energy mix by 2030 in order to show global leadership on renewables. In December 2018, the updated Renewable Energy Directive (2018/ 2001/EU), which includes this commitment, became effective.

- **Energy Efficiency**
One of the main goals of the package is to prioritize energy efficiency because doing so would reduce greenhouse gas emissions and benefit customers financially. Consequently, the EU has established legally obligatory goals to increase energy efficiency over present levels by at least 32.5% by 2030. This objective is stated in the Energy Efficiency Directive ((EU) 2018/2002), which has been in effect since December 2018.

- **Governance Regulation**
A strong governance framework for the EU's aim to substantially alter Europe's energy sector is part of the package. Each EU countries must create integrated 10-year National Energy and Climate Plans (NECPs) for 2021–30 in accordance with this policy. The NECPs describe how EU nations will accomplish their own goals for each of the five components of the energy union, including a longer-term perspective toward 2050. Since December 2018, the applicable law, the Regulation on the Governance of the Energy Union and Climate Action (EU) 2018/1999, has been in effect.

- **Electricity Market Design**
Another component of the package is to create a contemporary design for the European power market that is more adaptable, more market-based, and better positioned to include a larger share of renewable energy sources. The four strands of the electrical market design elements include two new energy laws, one on risk mitigation, and one detailing a more significant role for the Agency for the Cooperation of Energy Regulators (ACER).

Number of initiatives are raised in order to help with reaching the climate goals. Even in the **Sustainable Development Goals** (SDGs) up to three points are related to energy and climate problems, which is shown on the Figure 1. The SDGs are a collection of 17 interlinked global goals designed to be a *"shared blueprint for peace and prosperity for people and the planet, now and into the future"* [3].

Although there is still more work to be done to combat climate change, low-carbon technologies and new markets have already emerged in the years after the Paris Agreement came into force. A growing number of nations, regions, cities, and businesses are setting carbon neutrality goals. In all economic sectors, which account for 25% of emissions, zero-carbon solutions are becoming more competitive. The electricity and transportation industries are where this trend is most obvious, and it has given early adopters numerous brand-new business prospects.

Figure 1: Sustainable Development Goals Related to Climate and Energy [3]

Zero-carbon alternatives may be competitive by 2030 in industries responsible for more than 70% of world emissions.

3.2 The Photovoltaic Market

Nowadays, renewable energies are becoming more important in the generation of electricity. Fossil fuels do not provide a sustainable option for the future, since they are non-renewable sources of energy that cause environmental harm.

Solar energy is currently the most popular energy source since it is noiseless and clean. The aforementioned features make photovoltaic technology one of the most desirable renewable technologies. According to some studies, the world has very recently installed enough solar panels to generate 1TW of electricity directly from the sun.

The largest contributor to this capacity is China, who broke through 100GW in late 2016. The European Union hit 100GW in 2015, just before China. In the United States, the 100GW number was hit in the first quarter of 2021. These three regions represent more than half of the world's installed solar capacity [4].

Other than improving efficiency and reducing CO2 emissions to the atmosphere, this technique offers further benefits. It is a standard technology that can be used by any user, whether domestic or industrial, and can be installed in practically any area with a minimal number of hours of sunlight. Its energy source, the sun, is inexhaustible and free. As a result, it becomes the best tool for creating a large-scale centralized generating system [5].

ETSEIB

3.2.1 Energy Performance of Buildings Directive

The building sector is essential to achieving the energy and environmental objectives of the EU. Better and more energy-efficient constructions will simultaneously raise living standards for residents, reduce energy poverty, and have a positive impact on the economy, society, and health through increased indoor comfort and green jobs.
To boost the energy performance of buildings, the EU has established a legislative framework that includes the **Energy Performance of Buildings Directive 2010/31/EU** and the **Energy Efficiency Directive 2012/27/EU** [6].

Together, the directives promote policies that will help:

- achieve a highly energy efficient and decarbonized building stock by 2050
- create a stable environment for investment decisions
- enable consumers and businesses to make more informed choices to save energy and money

The aforementioned directive has been renewed in the well-known **"Winter Package"** [7]. The Energy Performance of Buildings Directive requires that all 28 EU member states had to ensure that all new buildings were nearly zero-energy by the end of 2020 while all new public buildings had to be nearly zero-energy after 31 December 2018. The term "nearly zero-emission building" (NZEBs) describes a building with extremely high energy efficiency, and the nearly zero or very low amount of energy required should be entirely or largely met by renewable energy, including energy generated on-site or nearby [8].

According to the **Global Market Outlook For Solar Power 2022-2026** [9], which is the most authoritative market analysis report for the global solar power sector, there are 4 main takeaways worth highlighting.

1. **Global solar reaches Terawatt level**
   As mentioned before, the global solar fleet reached a Terawatt capacity in April 2022 after doubling in size in just three years from 2018. The world market for solar energy is expanding rapidly. From 100 GW in 2012, it took around ten years for the global solar capacity to reach 1 TW. According to SolarPower Europe, the worldwide solar market will more than double to 2.3 TW in 2025 in only 3 years.

2. **Solar: the fastest growing energy technology**
   With more than half of the 302 GW of renewable energy deployed globally in 2021, solar energy is still the renewable energy source with the highest growth. With 168 GW of additions, solar installed over 70 GW more than the next greatest installer – wind – and more than all non-solar renewables combined.

3. **Global solar shares**
   China maintained its global leadership in 2021, installing twice as much solar power capacity as the United States, the second-largest market, with a 14% an-

nual growth rate and an all-time high of 54.9 GW of new solar. Despite this, the United States had tremendous growth, with 42% more solar energy installed in 2021 than in 2020. With 14.2 GW of solar installations, India recaptured third place.

4. **European solar trajectory**
The Europe area maintained its positivesolar trajectory, adding 31.8 GW of new solar capacity, which represents a 33% increase and is notable for being only 0.1 GW from Outlook's predictions for 2021. In addition to the EU climate goals, the effects of the Russian war in Ukraine and the challenges associated with energy security are what is driving the continent's transition toward renewable energy, with 25 of the 27 EU member states intending to install more solar in 2022 than in 2021.

3.3 Aim and Motivation of the Thesis

The main motivation behind this thesis was the ongoing European **ePLANET** project which is being implemented under the **Horizon 2020** initiative. The objectives of the project are described in detail in the section below. The Horizon 2020 is the EU's research and innovation funding programme from 2014-2020 with a budget of nearly €80 billion, which is the largest programme in the history of EU [10]. H2020 supports financially the Innovation Union, a Europe 2020 flagship initiative focused on ensuring Europe's global competitiveness. The main objective of H2020 is to enhance the EU's position as a global leader in science, innovation, and technology. This would help Europe become more appealing for investments in research and innovation.

The thesis provides a theoretical review of pieces of literature regarding amongst others Spanish regulation concerning photovoltaic self-consumption, energy communities, GIS projects, and a state of art of similar projects. Additionally, it explains the purpose of the European project according to which it was developed, and the achievements and purpose of the company that was supervising the thesis.

The experimental part is aiming to characterize the buildings energy consumption based on real data in six municipalities in Spain, in order to later cross it with the potential solar PV generation and highlight the possible areas where the potential energy communities could be created. The latter one was not achieved in the duration of this thesis. However, the findings of this thesis make it a perfect base for future work and further development. The main focus of the experiment of this work is data harmonization, unification and aggregation, in order to extract daily load profiles of particular addresses. Daily load profiles are crucial for analyzing the potential membership in an energy community after comparing it with the solar production of individual buildings.

The tool used to develop the experimental part of this thesis is a script developed in a Python environment. The script is scalable and universal, which means it can be run on consumption datasets of different locations, not only the ones involved in this project.

Regarding this, the novelty of the project was the fact that it is based on real data, it aggregates dataframes with different timestamp frequencies, and it is involved in the real ongoing European project, i which the findings and the outputs of this thesis will be used.

3.3.1 ePLANET

The ability of local and regional authorities to successfully coordinate is crucial to the transition from fossil fuel-based energy systems to renewable energy sources. The EU-founded ePLANET project promotes the digitalization of measures and plans, providing an interoperable ecosystem of data and tools supporting energy transition decision-making, with the assistance of innovative multi-level clustering governance within the public sector. The project makes use of the most innovative tools in big data and artificial intelligence to utilize and exploit the available energy data to boost energy transition plans. The part of the promoting leaflet is presented on the Figure 2 [11]. Before focusing on extending these tools and approaches across Europe, the project will first develop a proof of concept in regional pilots:

- Girona, Spain
- Crete, Greece
- Zlin, Czech Republic

Figure 2: ePLANET: Logo and Partners
[12]

The project has a duration of 3 years, it started on the 1st of September 2021 and its end date is scheduled on the 31st of August 2024. It obtained almost 1.5 million Euros funding from the European Union under the *Societal Challenges* funds regarding secure clean and efficient energy. It is being coordinated by the **Centre Internacional de Metodes Numerics en Enginyeria (CIMNE)** - a research center in Catalonia, Spain.

The main objectives of the ePLANET projects are:

1. To build a transparent and consistent platform for information exchange that enables direct comparison of Energy Transition Measures (ETM), policies, and data, amongst public agencies, as well as systematic data integration.

2. Establishing multi-level working groups with relevant public agencies and stakeholders to facilitate information exchange and a team-based approach in the creation and revision of Energy Transition plans.

3. To enable digitalization of the energy transition plans and energy transition measures of the public authorities as base for coherence, sharing, coordination and integration.

4. To strengthen current networks and assist new adopters.

5. To formalize multi-level working groups and ePLANET networks.

Figure 3: ePLANET: Main Goals
[12]

An impact the of ePLANET project
The ePLANET project will have a significant impact since it seeks to participate, cooperate, and process data from most energy-transition-related initiatives in numerous parts of Europe during the next few years. However, it is also challenging to predict how large this influence will actually be. The project's effect will be determined by how well each of the main indicators aligns with the goals set by the European Commission on the Horizon 2030, which must be met by the various regions.

According to the official ePLANET's website [12] the agreed actions are expected to have the following impacts amongst others:

- 1 harmonized set of Energy Transition Measures and policies across local, regional and national authorities
- 1 Common Data Model as a reference for data sharing
- 764 stakeholders engaged in ePLANET governance clusters (national, regional, local), including Public Authorities and private stakeholders
- 32 policies influenced
- 85 investments in sustainable energy
- 177,6 GWh/year primary energy savings

- 12,26 GWh/year renewable energy production

ePLANET Platform Specifications
The ePLANET platform's objective is to give municipalities digital assistance with the planning, drafting, and monitoring of their sustainable energy transition plans. In further detail, the platform will enable users to:

- Monitor the progress and effectiveness of existing SECAPS
- Consult and contribute to a database of energy transition measures, associated investment, and avoided energy/emissions
- Access data visualizations at a geographical level that can support municipalities and regions in drafting their energy transition plans

In order to provide these services the platform will be structured in three main modules:

1. Energy transition plan monitoring system
2. Public energy transition measures database
3. GIS-based support tool for energy transition planning

This thesis's main focus is put on the last one related to the Geographical Information System-based tool. This module will offer GIS-mapped representations of a variety of data that may be useful for developing a municipal energy transition strategy. The platform will establish connections with the relevant data sources, carry out harmonization tasks, and enable the visualization of a number of layers of data. Users will also be able to make benchmarking comparisons between regions of interest and shift between several graphical scales - building block, municipality, and postal code levels. One example of the usage of a GIS-based support system may be to match energy consumption data at the building block level with PV rooftop potential studies in order to design collective self-consumption projects for energy communities.

This is exactly the topic of this thesis - collecting energy and gas consumption for all buildings of the pilot municipalities in Girona and matching it with the Spanish cadastre thanks to the harmonization script which will be developed throughout this thesis. Through the cadastral data it is possible to obtain the exact geographical coordinates of a building block, and therefore visualize data related to that block on the map.

The methodology is explained in the sections below. Additionally, the comparison of all platforms with a similar concept or comparable purpose will be presented in the *State of the Art* section.

3.3.2 CIMNE - BEE Group

As mentioned above, this thesis was developed under the supervision of the Catalan research center **CIMNE - the International Centre for Numerical Methods in Engi-**

neering, in an autonomous department called **BEE Group - Building Energy and Environment Group**. As a joint venture between the Government of Catalonia and Universitat Politècnica de Catalunya (UPC), in conjunction with UNESCO, CIMNE was established in 1987 at the core of this prestigious technical university. The development of numerical methods and computational procedures is the goal of CIMNE in order to advance knowledge and technology in applied sciences and engineering [13].

Figure 4: BeeGroup - Building Energy and Environment Group
[14]

The BeeGroup name stands for Building Energy and Environment. The main focus of this group is on big data and artificial intelligence to improve the use of energy in buildings and the environment. The research and development efforts are concentrated on creating better building energy management through increased accuracy, quicker reaction, the implementation of adaptive and predictive control, and increased responsiveness of buildings to user need and desires. Additionally, its focus is on making energy data more cost-effective, applicable, and reliable for businesses and professionals by utilizing big data analytics, personalized energy services, adaptable visual interfaces, and mobile apps. Moreover, the goal is to do all of the things mentioned above in a customer-friendly way by choosing the most relevant information, improving control, and customizing the experience [14].

The main research lines are:

- Big data Analytics for Energy Efficiency in Buildings
- Energy Communities & Energy Positive Municipalities
- Demand Response
- Energy Empowerment & Users Behavior
- Biodigesters
- Energy Positive Buildings

4 Energy Communities

Energy communities plan communal, citizen-driven energy initiatives that advance the transition to renewable energy while elevating residents. They help to broaden public support for renewable energy initiatives and make it easier to attract private investment for the clean energy transition. By boosting energy efficiency, reducing consumers' power costs, and generating local jobs, they also have the potential to directly benefit residents. Energy communities can contribute to the flexibility of the electrical grid by encouraging public participation in demand response and storage.

An association, cooperative, partnership, non-profit, or other legal entity controlled by shareholders or local members, generally oriented to value rather than profitability, committed to distributed generation and realization activities of a distribution network manager, or supplier at the local level, can be referred to as an energy community [15].

"Energy communities are one of the key elements for achieving the EU's energy transition: by 2050, half of Europe's citizens could be producing up to half of the EU's renewable energy." [16]

Energy communities have a number of advantages, including increasing the share of renewable energy sources in the electricity supply, reducing power losses while delaying future investments in transmission and distribution infrastructure, diversifying centralized energy investments, and increasing the number of actors who benefit from the electricity generation activity.

On the other hand, it also brings some drawbacks which can occur in the event of high PV generation and low electricity demand which might cause reverse power flows at a low voltage grid. Also, the cost per kW of installation of PVSC systems is more expensive than the installation of the large-scale PV installations [17].

According to European law, the *Clean Energy Package* recognizes some types of community energy initiatives as "energy communities". Energy communities may be seen as a form of "organizing" collective energy actions around transparent, democratic participation and governance, and the supply of benefits for the members of the local community.

There are two official definitions of energy communities [18]:

1. **Citizen Energy Community (CEC)**
   Included in the revised *Internal Electricity Market Directive (EU) 2019/944*. It is described as a legal body that is based on voluntary and open participation, and its effectively governed by stakeholders or members who are neutral people, local government entities, including municipalities, small businesses, and micro-enterprises.

2. **Renewable Energy Community (REC)**
   Included in the revised *Renewable Energy Directive (EU) 2018/2001*. It is a legal entity that, in accordance with the relevant national law, is based on open and vol-

untary participation, autonomous, and effectively controlled by shareholders or members who are situated close to the renewable energy projects that are owned and developed by that legal entity; the shareholders or members may be individuals, SMEs, or local authorities, including municipalities.

The creation of energy communities is a challenge. However, Europe is very invested in promoting this concept and in decarbonizing the environment. Now, in the EU there are over 3500 renewable energy cooperatives owned by citizens: the vast majority are in the Netherlands, the UK and mainly Germany [19]. It might be surprising since Germany has three times as much installed solar power as Spain, even though it had about 1896 hours of sunshine in 2020, compared with almost 3000 hours in Spain [20]. However Spain, since the change of regulations, is putting emphasis on the development of ECs. The perfect example is the leading energy cooperative **Som Energia** which has over 75,000 members across Spain.

Energy communities are developing at such a strong pace that the European Commission came up with an idea in 2022 of creating a **Energy Community Repository** on its official website, which provides technical assistance, legal analysis, best practices, and tools to boost energy communities across Europe [21].

The key challenge, highlighted by Aura Caramizaru and Andreas Uihlein in *Energy communities: an overview of energy and social innovation* report, which has to be faced is to ensure the cost-efficiency of ECs beyond locally-derived benefits. The reduced grid fees are only beneficial for the members of a community, meaning that it will transform into the cost for other grid users, outside the EC [19].

4.1 Spanish Regulatory Framework

Between 2015 and 2018 (RD 900/2015, 2015), Spain had some of the strictest photovoltaic self-consumption (PVSC) laws in the whole world. By neglecting any remuneration for the excess electricity exported to the grid for residential prosumers on the one hand, and by imposing a backup charge on the self-consumed electricity for commercial and industrial segments on the other, this regulation prevented the economic viability of PVSC installations.

The new regulation, which enhances the financial conditions of prosumers and simplifies administrative processes, was enacted in October 2018 (Royal Decreto-Ley - **RD-L 15/2018**, 2018) and further improved in April 2019 (**RD 244/2019**) [22]. These adjustments improved PVSC installations' profitability and may have contributed to the establishment of new business models.

The current regulation implements beneficial for potential prosumers conditions [23]:

- The backup charge on the self-consumed electricity removed
- The surplus electricity exported to the grid is remunerated

- Shared self-consumption allowed

The new rule generally makes things better for PV prosumers. On the one hand, it makes the administrative and technological requirements simpler. On the other hand, by eliminating the backup tax and recognizing prosumers' right to sell any excess power to the grid, it increases the financial feasibility of PV systems.

These principles are important because they expressly recognize the right to self- consumption and the potential of shared self-consumption between several end-consumers. The extra power exported to the grid shall be handled "under the same conditions as any other production site," according to this law (art. 5 Title II). Allowing shared self-consumption can increase PVSC in urban areas and help maximize the share of electricity self-consumed over the total produced, minimizing the impact of PVSC on the electricity system. Removing the backup charge increases the economic incentive for self-consumption, compensating the surplus electricity reduces the incentive for going off-grid and allowing shared self-consumption can increase PVSC in rural areas. Finally, the simplification of administrative procedures reduces related soft costs and information barriers likely to hinder the expansion of residential installations. All these measures improve the profitability of PVSC projects which in the sense of Energy Communities is very important.

Prosumers may now select between a **general system** where they can directly sell the excess power they generate to the grid under the same terms as any other producer or a **simplified monthly net billing** process where the profits are directly discounted off the electricity bill. The total profits are constrained by the cost of the energy portion of the electricity bill, the monthly billing, and the maximum installed capacity of 100 kWp, whereas the net billing system offers higher profitability because profits are realized as savings and are therefore not taxable, and the surplus electricity exported to the grid is exempt from the generation tax and the grid-access charge. On the opposite, the general system would yield lower profitability but total profits are not constrained by any of the aforementioned conditions.

There are some crucial limitations concerning the self consumption. According to the **Royal Decree-Law 244/2019** which regulates self-consumption, the generation unit does not necessarily have to be in the building where self-consumption takes place, it can instead be a *"nearby facility"*. However, the regulation poses a limit of **500 meters** between the RES generation units and the self-consumption point, and both shall be connected at **low voltage**. This is a barrier since it can limit more projects and initiatives especially for collective self-consumption and for Renewable Energy Communities. Additionally, current rules in Spain have established a power limit between installations that can only be used for self-consumption and those that can also serve as power producers and sell all of their generation on the wholesale market. This limit has been fixed at 100 KW, thus any facilities operating below this threshold may only be used for self-consumption.

One of the newest regulation that was introduced on January 2022 allows a mode of self-consumption based on a static sharing of the generated energy. This means that established coefficients must be precisely recognized in order to allocate all of the generated energy among the participants of the local energy community. This results in a more effective distribution. The theory underlying this improved efficiency is the rise in the probability of matching the produced energy to the aggregated demand of the associated customers.

5 State of the Art

There exist a number of platforms and companies regarding the current development of Energy Communities and actions related to this, which either promote and help the creation of ECs, simulate them or monitor the energy usage across a given area.

As mentioned in previous sections, the main topic of the ePLANET project, as well as this thesis, is to develop a tool that would help the creation of a bigger GIS-based platform that would visualize the energy and gas consumption of all related buildings together with their solar PV rooftop potential in order to design collective self-consumption projects for energy communities. In this section, there will be presented a list of already existing platforms and companies whose activities are focused on the same topic - Energy Communities.

The majority of presented platforms are tools to map solar PV potential, irradiation, or energy consumption, while some of them are open-source which makes it extendable for other areas. The interesting feature is that part of them are combined with deep learning algorithms (such as CNN - Convolutional Neural Network), Agent-Based Modelling (ABM) or Decision Support System (DSS) depending on the purpose. The data is usually retrieved from public authorities in most cases. The ones that are GIS-based most frequently use ArcGIS - the web-based mapping software and interactive map builder.

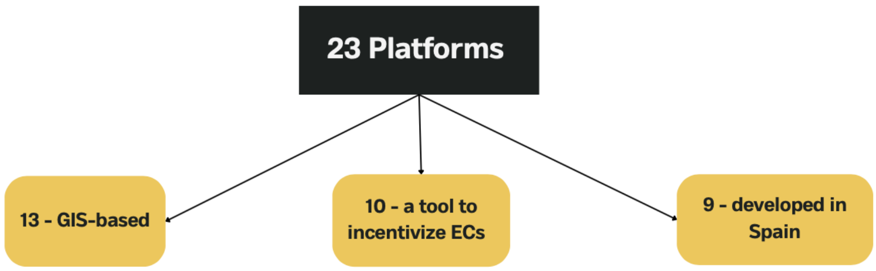

Figure 5: Analyzed Platforms structure

As it is shown on the Figure 5, this research analyzed 23 different platforms and projects which are directly related to the topic of ePLANET. The scope included among others: web service platforms, software, and consultancy. 13 platforms are GIS-based, while the rest is a interactive tool to incentive municipalities for energy communities creation. 9 out of 23 platforms subjected to this analysis were developed in Spain.

5.1 GIS-Platforms

The most relevant fraction of all the platforms are the one that are GIS-based, since it is the key characteristic of the potential ePLANET platform that is being developed.

1. **Som Comunitat Energetica**
   Developed in Institute for Energy Research of Catalonia (IREC) in Spain. The

Figure 6: Som Comunitat Energetica Logo
[24]

main objective is to boost the energy transition of the building stock towards decarbonization by promoting the creation of energy communities from an inclusive and supportive perspective. It is an interactive map based on the satellite picture which includes an algorithm for recognizing potential energy communities. The main feature is the information about the solar radiation of each rooftop in the neighborhood and ability to simulate the creation of EC in order to discover the energy and economic savings.

2. **Grid Singularity**

Figure 7: Grid Singularity Logo
[25]

Energy exchange software development company that simulates and operates interconnected grid-aware energy marketplaces. Grid Singularity facilitates market participation by connecting aggregators and grid operators through an application interface. Grid Singularity Exchange provides the software tools, enabling prosumers and consumers organized in energy communities, with the support of local aggregators, to simulate and implement peer-to-peer and community energy trading through the creation of active local energy markets. It is a *"digital twin"* representation of physical energy systems and energy markets.

3. **Project Sunroof**

Figure 8: Project Sunroof Logo
[26]

Developed by the collaboration between Google, E.ON, and Tetraeder.solar, which is a solar potential map based on Google Maps and machine learning processes, and allows to design of a future photovoltaic installation and expected reduction of the electricity bill. It takes into consideration shadows, all possible sun positions and historical cloud, and temperature patterns.

4. **AURORA - Smart Solar Power**
A platform that creates an entire engineering design and sales proposal with just

Figure 9: AURORA - Smart Solar Power Logo
[27]

an electric bill and an address after the selection of a particular rooftop. It uses LiDAR-assisted modeling and conducts shading analysis. After this process, it is able to calculate the price and simulate the energy production of the potential installation. It operates in the USA.

5. **Roofpedia**
An automatic mapping of green and solar roofs for an open roofscape registry and evaluation of urban sustainability. It combines deep-learning and geo-spatial techniques demonstrating the feasibility of an automated methodology that generalizes successfully across cities with an accuracy of detecting sustainable roof of up to 100% in some cities. The method identifies both green and solar roofs, which adds another dimension to understanding the sustainability of the roofscape across cities. The model is created with international scalability in mind and

Figure 10: Roofpedia Logo
[28]

it can be used in multiple cities around the world with different urban morphology and building typology.

5.2 Tools for the municipalities for Energy Communities creation

A number of companies provide a variety of tools that helps municipalities or citizens implement the Energy Communities concept. From just providing consultancy services, through sharing relevant consumption or potential solar generation data, to platforms through which it is possible to sell the excess electricity. Below a few of them are explained.

1. **Tetraeder.solar**
   Big data for sales and power grid planning. The company helps with utilization

   Figure 11: Tetraeder.solar Logo
   [29]

   of huge amounts of data, such as a solar potential analysis of entire areas or even countries. The focus is on the the development of forecasting and planning tools in the field of renewable energies, especially in the fields of photovoltaics, solar thermal energy and the use of energy storage, as well as expansion of the PV and charging infrastructure. The services are mainly used by municipalities and public utilities / energy supply companies to support public relations and sales.

2. **KM0**
   The company promotes, designs, implements and manages the Energy Communities, accompanying town councils and citizens throughout the energy transition process, empowering them and involving them in a new way of doing things. One of the goals aims to implement 1GW of decentralized renewable energy in Catalonia, which is equivalent to a nuclear power plant. The generation plants that the company supports have citizen financing since the energy transition must have its engine in a social transformation, which allows citizens to have ownership of electricity generation, and move from a centralized system in the hands of large institutions, to a decentralized citizen-based system.

Figure 12: KM0 Logo
[30]

3. **Power Quartier**
   The app accompanies households and businesses in the energy transition, provid-

Figure 13: Power Quartier Logo
[31]

   ing insights into their consumption, and showing them how photovoltaics reduce energy bills. The flagship service is a "Neighborhood Power" exchange, allowing an electricity-producing household to sell any surpluses to surrounding households. Both, buyers and sellers participate in the local energy community via the PowerQuartier app that shows energy data and generation sources. For example, on sunny days, users can see that the local photovoltaic system covers the whole community's electricity demand. On rainy days, you see how much electricity the regional energy supplier is delivering to the community. The app supports municipalities in turning their political ambitions regarding the energy transition into reality and showing the population possible courses of action.

4. **SimStadt**
   The aim is to provide a simulation environment that couples building require-

Figure 14: SimStadt Logo
[32]

   ment analyzes with decentralized renewable feed-in via grid simulations and thus enables scenarios for load management, storage dimensioning and demand developments to be calculated. In the current stage of expansion, SimStadt is able to use data from a real urban development situation or a planning status for energy analyzes of buildings, urban districts, entire cities and even regions. The application scenarios range from the high-resolution simulations of the building heating requirement to potential studies for photovoltaics to the simulation of building renovation and renewable energy supply scenarios.

5. **EnerMaps**
It is a European project, that was also involved in the H2020 projects, which pro-

Figure 15: EnerMaps Logo
[33]

vides an open data management tool that improves data management and accessibility in the field of the renewable energy industry. It aims to collect centralized energy data in a common platform with long-term support.

All the above-mentioned platforms and services have similar features to the upcoming ePLANET platform. It is important to analyze all the already existing solutions before developing a new idea. The closest concept to the ePLANET one is the Som Comunitat Energetica service since it is also GIS-based and it is possible to simulate a potential energy community by simply choosing the rooftops in the proximity of a 500m circle.

However, contrary to the ePLANET project, it does not include real and detailed energy consumption data on a building level - the calculations are based only on the rooftop solar potential. This feature coupled with geographical coordinates and PV solar potential for individual rooftops makes the ePLANET GIS-based platform stand out from its competitors. With these characteristics, the potential creation of Energy Communities is more supported, because it is based on real historical data, and therefore it is more reliable. The Som Comunitat Energetica idea is missing the support for the local authorities and it's not involving municipalities in the creation of the ECs. In the case of the ePLANET project, it is crucial to use public buildings in the analysis and choose rooftops around with the most beneficial and varied load profiles in order to use as much generated electricity as possible.

In the later sections, the particular business case will be described and the whole methodology which was followed during the development of this thesis.

6 Particular Business Case

6.1 Overview of the Studied Case

Girona province, as one of three pilot areas involved in the aforementioned ePLANET project, is the main subject of this thesis. It is a province of Spain located in the north-eastern part of the autonomous community of Catalonia. Girona consists of 222 municipalities, out of which the main focus of this thesis is on six of them [34].

This particular province has been very active and energy conscious for a long time now. In November of 2008, the city of Girona started to participate in the Covenant of Mayors project which was established in the same year by the European Commission [35]. The goal of the initiative is to engage and support mayors to commit to reaching the EU's climate and energy targets. The Covenant of Mayors Signatories commits to:

- reduce greenhouse gas emissions on their territory,
- boost resilience and prepare for the negative effects of climate change,
- address energy poverty as one crucial key action to ensure an equitable transition.

In addition to pioneering a bottom-up strategy for tackling energy and climate change, the program soon outperformed expectations. With the backing of specialized offices and the strengths of a global multi-stakeholder movement, the project currently unites more than 9000 local and regional authorities from 57 different countries [36].

The parties to the Covenant agree to use an integrated strategy for both reducing emissions and adapting for them. Within the first two years after adhesion, they must create a Sustainable Energy and Climate Action Plan (SECAPs) with the objective of reducing CO2 emissions by at least 40% by 2030 and boosting climate change resilience. A signatory to the Covenant of Mayors is required to submit a Monitoring Report every second year after the adoption of its Action Plan regarding the mitigation and adaptation of these planned actions. Additionally, at least every four years the emissions inventory of the action agreements has to be updated in order to see the progress related to resilience to climate change, reduction of emissions, and energy consumption [37].

The number of municipalities participating in the Covenant of Mayors in the Girona Province increased from three in 2010 to over 190 in 2013 in only three years. Given that the municipalities collectively have a population of 700,000 inhabitants (state for the year 2013), the initiatives started by the project will benefit 93% of the province's citizens [38]. It is the outcome of the Girona provincial administration's efforts and involvement after signing the agreement. The perfect example of this is the fact that since 2012 the Province Council is Covenant of Mayor Territorial Coordinator [39]. Due to this, a specialized sustainable energy team has been created to support municipalities' mobilize investments in sustainable energy through innovative organizational and financial schemes.

The previously mentioned six municipalities of Girona that constitute the major subject of this thesis are listed below with their respective municipality codes:

1. Campdevanol (170366)
2. Cassa de la Selva (170448)
3. Palafrugell (171175)
4. Ripoll (171479)
5. Sant Feliu de Guixols (171609)
6. Sarria de Ter (171864)

The more detailed information can be found in the Table 1:

Municipality	Code	Population (2021)	Area [km^2]	Density [hab./km^2]
Campdevanol	173660	3 225	32.62	98.9
Cassa de la Selva	170448	10 505	45.21	232.4
Palafrugell	171175	23 046	26.88	857.4
Ripoll	171479	10 721	73.71	145.4
Sant Feliu de Guixols	171609	22 210	16.23	1 368.5
Sarria de Ter	171864	5 231	4.16	1 257.5

Table 1: Characteristics of the considered municipalities [40]

These are the cases for which data on gas and electricity consumption was made available by the public authorities. Within these municipalities there are examples of different energy sectors, among others: residential, industrial, offices, agriculture, or public spaces, making it the ideal illustration for this analysis.

6.2 Methodology

This thesis mainly focuses on the data analysis which is a method for collecting, transforming, and organizing data in order to make predictions about the future and well-informed data-driven decisions. One of the main programming environment used for the data science is **Python**. As the official Python website states: *"Python is an interpreted, interactive, object-oriented programming language. It incorporates modules, exceptions, dynamic typing, very high level dynamic data types, and classes. It supports multiple programming paradigms beyond object-oriented programming, such as procedural and functional programming. Python combines remarkable power with very clear syntax. Finally, Python is portable: it runs on many Unix variants including Linux and macOS, and on Windows"* [41]. Python includes many helpful libraries. The term *"Python library"* refers to a group of connected modules. It has collections of code that may be utilized repeatedly in different programs. For the programmers, it simplifies and makes Python programming more practical.

The library extensively used during the development of this thesis is the **Pandas library**. For data scientists, the Pandas library is fundamental. It is a machine learning package that offers a selection of analytical tools and configurable high-level data structures. It makes data analysis, manipulation and cleaning easier. Sorting, re-indexing, iteration, concatenation, data conversion, visualizations, aggregations, and other operations are supported by Pandas [42].

Second package frequently used in this analysis was the **Numpy library** which stands for *"Numerical Python"*. Large matrices and multi-dimensional data are supported by this well-liked machine learning package. It has built-in mathematical functions for quick calculations. One of the main characteristics of this library is the Array Interface [42].

Another library widely utilized in this study was the **Matplotlib library**. Plotting numerical data is the responsibility of this library. It plots highly specified images like pie charts, histograms, scatterplots, graphs, etc., and it is an open-source library [42].

It is vital to note that this methodology is crucial in highlighting the project's standout features.

6.2.1 Data Collection

The first step, and one of the most important one is the data collection. It is the process of gathering data for use in business decision-making, strategic planning, research and other purposes. It is an essential component of research initiatives and data analytics applications. Depending on how much information is required, data can be gathered from one or many sources. The following are some difficulties frequently encountered when gathering data:

- Data quality issues - raw data usually contains errors, inconsistencies, and other issues, it typically needs to be put through data profiling to identify issues and data cleansing to fix them.

- Finding relevant data - for the data scientists acquiring data to analyze may be a challenging tasks due to the variety of systems they must navigate. Data identification and access are made simpler with the usage of data curation techniques. For instance, building a data catalog and searchable indexes may fall under this strategy.

- Deciding what data to collect - fundamental issue which has a big influence on process time, expenses and complexity. When done incorrectly may cause a huge data loss.

- Merging large data - large amounts of structured and unstructured, and semistructured data are frequently present in a big data contexts. Because of this, the earliest phases of data collecting and processing are more difficult.

Data used for this thesis was gathered from the six municipalities in Girona which were mentioned in the previous section. The following sets of raw data were collected:

1. **Monthly electricity consumption**
   Obtained from the Distribution System Operator (DSO) - **ENDESA** which is a Spanish biggest electric utility company [43]. Data includes the period from January 2018 to December 2020, and the consumption is collected on the building level. It is worth mentioning that due to confidentiality and data protection, the DSO can share only the data of the building containing more than 4 clients, in order not to be able to extract the demand of an individual customer.

2. **Yearly gas consumption**
   Shared by Nedgia [44] - a leader in natural gas distribution in Spain. Data was collected on the building level manner for the years 2018 - 2021.

3. **Aggregated hourly electricity consumption**
   Downloaded from the DataDis platform [45] which collects electricity data available after request. Data aggregated on a postal code level for the time period: January 2018 - December 2020.

4. **Cadastral Data**
   Georeferenced data publicly available, contains official, legal documentation concerning the quantity, dimensions, locations, coordinates, and many others of individual parcels [46].

6.2.2 Data Preprocessing

Data preprocessing refers to the procedures that must be followed to alter or encode data so that the computer can quickly and readily decode it. Due to their varied origin, the majority of real-world datasets used for learning algorithms are very likely to have missing data, inconsistent results, and noise. Datasets would not produce high-quality results when applied to this noisy data because they would be unable to successfully find patterns. Therefore, data processing is crucial to raising the general level of data quality. Missing or duplicate numbers might present an inaccurate picture of the data's overall statistics. False predictions are frequently the result of outliers and inconsistent data points disrupting the model's overall learning process.

As some studies state, there are 4 steps in Data Preprocessing [47]:

1. Data Cleaning and Harmonization - filling in missing values, smoothing noisy data, resolving inconsistencies, removing outliers are all included in this step.

2. Data Integration - is employed to combine data from several sources into one bigger data repository.

3. Data Transformation - consolidation of the quality of data into alternate forms by

changing the value, structure, or format of data using data transformation strategies.

4. Data Reduction - decreasing the size of dataset in order to make it easier to handle by data analysis and data mining algorithms.

These steps were used during the development of this thesis. Especially the data harmonization step was the most time-consuming, since raw datasets had different origins. The objective was to have a structure that is common to all of the dataframes in order to apply the forthcoming procedures, which required executing a number of adjustments and unification processes to successfully combine all the various datasets. Operations implemented on each of the databases are explained below.

Firstly, the main goal of the data harmonization was to merge gas and electricity consumption data with the cadastral data, in an attempt to later visualize it on the GIS-based technology.

It is important to note that owing to confidentiality the real addresses were substituted by the fictitious names.

Monthly Electricity Consumption
The database obtained from the Spanish DSO - ENDESA consisted of the period of interest, type, and name of the street, as well as the number of the corresponding building, together with the number of clients assigned to it, a postal code, the name of the municipality, and 36 columns with the electricity consumption of each month during 3 years. At the beginning, the file included **18540** rows of the raw data. The primary structure is presented in Figure 16.

PERIODO	TP_VIA	DE_CALLE	NM_FINCA	CD_POSTAL	DE_MUNICIP	NUM_CLIENTES	CONSUMO_ENERO_2018	CONSUMO_FEBRERO_2018
ENE 2018 - DIC 2020	CALLE	CAMI HORTS	16	17124	PALAFRUGELL	1		
ENE 2018 - DIC 2020	CALLE	TRAMUNTANA	0	17124	PALAFRUGELL	1		
ENE 2018 - DIC 2020	PASAJE	ESTACIO	13	17124	PALAFRUGELL	1		
ENE 2018 - DIC 2020	PARAJE	DE SUBIRÁ	9	17124	PALAFRUGELL	1		
ENE 2018 - DIC 2020	CALLE	POUES	20	17124	PALAFRUGELL	1		

Figure 16: Monthly Electricity Consumption - Raw Dataset

A number of adjustments had to be performed in order to obtain the required structure of the dataframe. Firstly, all the column names were translated to English, and rows without values were deleted with the $df.dropna()$ function. These empty cells without values were the result of the previously mentioned confidentiality aspect which made it impossible to share the data of individual clients. It means that the electricity consumption was only given if the *NUM_CLIENTES* was larger than 4. This operation caused the dataset to shrink to **2408** rows which will be the number of the valid database for further processes. The next adjustment that was performed was the removal of the irrelevant columns for further work: *PERIODO*, *NM_FINCA*, *NUM_CLIENTES*. To make the forthcoming merge more detailed, an additional column with the corresponding municipality code was created to able to split the whole dataset into 6 smaller ones containing information from only one municipality. Furthermore, the columns with street names and numbers were joined together because the logic later required only one column as a parameter. The final structure is presented below in Figure 17.

CODE	CD_POSTAL	MUNICIPALITY	STREET	NUMBER	JANUARY 2018	FEBRUARY 2018	MARCH 2018	APRIL 2018	MAY 2018
170366	17530	CAMPDEVANOL	PAPER	0	57.133	51.135	34.339	34.359	35.777
170366	17530	CAMPDEVANOL	JUMP	1	1561.01	1476.0	1525.576	1377.736	1428.691
170366	17530	CAMPDEVANOL	JUMP	3	2049.971	1781.0	1848.729	1652.514	1616.759
170366	17530	CAMPDEVANOL	ELM	1	1890.991	1738.203	1857.177	1718.686	1656.104
170366	17530	CAMPDEVANOL	ELM	3	5021.181	4407.363	4433.939	3883.646	3574.038

Figure 17: Monthly Electricity Consumption - Final Structure

Yearly Gas Consumption

The file obtained from another Spanish DSO - Nedgia, included only yearly data, so the number of columns was way smaller than in the case of electricity consumption. On the contrary, it included data from more municipalities, than the ones considered in this thesis, which had to be removed. The number of raw data was equal to **12 636** rows. The dataframe was composed of the name of the province and the municipality, street and number, the number of clients for the corresponding address, and most importantly, the gas consumption of the years 2018 - 2021. It is shown on the Figure 18 and the final version on the Figure 19 below.

FINCA	MUNICIPI	Vol PS	Consum 2018	Consum 2019	Consum 2020	Consum 2021
GIRONA@AGULLANA@BAKER_STREET@6@	AGULLANA	4	997	987	892	1.015
GIRONA@AGULLANA@BAKER_STREET@8@	AGULLANA	6	1.473	1.852	1.280	1.253
GIRONA@AGULLANA@EVERGREEN_TERRACE@21@	AGULLANA	4	541	479	441	447
GIRONA@ALP@SESAME_STREET@0018@A	ALP	11	1.687	2.013	1.585	1.672
GIRONA@ALP@SESAME_STREET@0018@B	ALP	7	1.827	1.952	1.814	1.768

Figure 18: Yearly Gas Consumption - Raw Data

CODE	MUNICIPALITY	STREET	NUMBER	Consum 2018	Consum 2019	Consum 2020	Consum 2021
170366	CAMPDEVANOL	BEAUTIFUL	3	4862	4171	4345	3761
170366	CAMPDEVANOL	CORONATION	5	4595	3791	3808	3997
170366	CAMPDEVANOL	ELM	1	524	419	432	559
170366	CAMPDEVANOL	ELM	3	999	920	948	815

Figure 19: Yearly Gas Consumption - Final Structure

To organize it, only data from the 6 municipalities was extracted from the raw dataset, and the column with addresses was split into a few, based on the @ separator. Columns that were not relevant were deleted (*Vol PS*), instead of which, the one with the municipality code was added, using the *select*() function from the *Numpy* library. The function requires two parameters: the list of values (municipality codes), and conditions (names of municipalities) in order to return the required array which will be saved in the dataframe. After all the operations the final number of rows was **1790**.

Cadastral Dataset

The dataframe that includes all the cadastral information was the most crucial one since it linked two previous datasets together - electricity consumption and gas consumption. The main goal of this operation was to be able to represent all the data on the GIS-based platform. The geo-referenced information can is available for everyone to download from the Cadastral website and includes a lot of information about every parcel in the given municipality. The raw dataset is presented in Figure 20 and followed by the final, structured version in Figure 21.

gml_id	text	designator	xcoord	ycoord	cadastralReference	CODPOS	MUNDISSEC	CODIMUNI
17.040.87	VERNEDA	5	2.16702097931723	42.2319274463705	1360516DG3716S	17530	17036601003	170366
17.040.87	VERNEDA	7	2.1670883757805	42.2319354109677	1360515DG3716S	17530	17036601003	170366
17.040.87	VERNEDA	9	2.16714257597063	42.2319423790818	1360514DG3716S	17530	17036601003	170366
17.040.87	VERNEDA	11	2.16724575761197	42.2319569032742	1360513DG3716S	17530	17036601003	170366

addressWithoutDeletingStreetType	address	OBJECTID	MUNICIPI	NOM_MUNI	ESTAT_CONSERVACIO	ANY_CONSTRUCCIO	US_DOMINANT
CL VERNEDA 5	VERNEDA 5	741797	170366	Campdevànol	functional	1971	1_residential
CL VERNEDA 7	VERNEDA 7	741796	170366	Campdevànol	functional	1971	1_residential
CL VERNEDA 9	VERNEDA 9	741795	170366	Campdevànol	functional	1971	1_residential
CL VERNEDA 11	VERNEDA 11	741794	170366	Campdevànol	functional	1971	1_residential

NUM_IMMOBLES	NUM_VIVENDES	SUP_TOTAL_CONSTRUIDA_M2	SUP_GRAFICA_PLANTA_M2	FITXA
1	1	125	55	https://www1.sedecatastro.gob.es/CYCBienInmueble/OVCListaBienes.a
1	1	90	55	https://www1.sedecatastro.gob.es/CYCBienInmueble/OVCListaBienes.a
1	1	90	60	https://www1.sedecatastro.gob.es/CYCBienInmueble/OVCListaBienes.a
1	1	105	79	https://www1.sedecatastro.gob.es/CYCBienInmueble/OVCListaBienes.a

Figure 20: Cadastral Data - Raw Dataset

STREET	NUMBER	xcoord	ycoord	cadastralReference	CD_POSTAL	CODE	MUNICIPALITY	US_DOMINANT
VERNEDA	5	2.16702097931723	42.2319274463705	1360516DG3716S	17530.0	170366.0	CAMPDEVANOL	1_residential
VERNEDA	7	2.1670883757805	42.2319354109677	1360515DG3716S	17530.0	170366.0	CAMPDEVANOL	1_residential
VERNEDA	9	2.16714257597063	42.2319423790818	1360514DG3716S	17530.0	170366.0	CAMPDEVANOL	1_residential
VERNEDA	11	2.16724575761197	42.2319569032742	1360513DG3716S	17530.0	170366.0	CAMPDEVANOL	1_residential

Figure 21: Cadastral Data - Final Structure

The only relevant columns which contributed to the further development of this paper were: street name together with the number of the building, x and y coordinates, cadastral reference, postal code, municipality code, name of the municipality, dominant functionality of the real estate and a total surface area.

6.2.3 Data Harmonization

As previously mentioned, this step took the longest time to complete since the dataframes had different origins and their structures weren't compatible with each other. Different parties have different ways of spelling street names, different rules of using prefixes, abbreviations, punctuation, special characters, and many others. With these differences merging dataframes becomes very complicated. The main challenge during this step was to find as many matches between two datasets as possible in order to obtain a final database with individual addresses, their gas and electricity consumption, and geo-reference. The databases on which the merging was performed are: monthly electricity consumption with cadastral data, and yearly gas consumption, again with cadastral data.

The idea was to create three main datasets during the matching process: perfect matches, close matches, and problematic cases without a match. The latter one would shrink after each iteration of the code because of the extraction of new matches.

Merge Function

One of the Pandas' key features is its high-performance, in-memory join and merge operations. One-to-one, many-to-one, and many-to-many joints are among the types of joints that are implemented by the *pd.merge*() function. The result of the merge is a new dataframe that combines the information from the two inputs. Concerning this project the mentioned function took four different arguments: municipality, postal code, street name, and the number of the building. To make the merge more detailed, both consumption datasets were split into 6 smaller ones, each corresponding to a particular municipality. It was done in order to omit the problem of the same street name in different municipalities.

1. **Electricity Consumption & Cadastral Data Merge**
   During this step, one issue appeared with one municipality, where there was the same name of the street but a different street type: *Plaza Estacio* and *Avenida Estacio*. After noticing this special case, the street type was taken into account after the merging. However, in the cadastral dataframe the type of the street was only included in the *"addressWithoutDeletingStreetType"* column which consisted of the abbreviation of the type, and it was not compatible with the way of presenting the street type in the consumption data. Hence, a new Python file was written which replaced the abbreviation of the street type with the full description, as in the example code below:

   ```
   df['tp'] = df['tp'].str.replace('PZ', 'PLAZA')
   ```

   After the merging function, if the street name was the same, the code compared the types of the street and deleted the cases where it was different.
   The raw database consisted of 2409 rows, as stated before. After performing all these operations, the program found **1052** perfect matches.

2. **Gas Consumption & Cadastral Data Merge**
   The same logic was applied in the case of gas consumption. However, the case of the street types was not valid because the raw dataset did not distinguish between types of streets. In this case, one street type was assumed after some investigations. After all, the number of matches was **1034** out of a primary database.

Fuzzy Joint

After the merge algorithm where only perfect matches were found, the remaining addresses had to be assigned to their corresponding ones in different sets of data. The main challenge of this step was the fact that different datasets had the different spellings of the street names. A few examples included using special Catalan characters like the letter ç or using abbreviations like *"St"* for *"Sant"*. A unique Python tool created for solving these kinds of problems called Fuzzy Merge was applied.

It can be downloaded and installed from the *fuzzywuzzy* library. It is a smart data preparation feature that is used to apply fuzzy matching algorithms when comparing columns in order to find matches across the tables that are being merged. Contrary to the previous method, fuzzy logic can take only one parameter, meaning one column. Therefore, the columns that have been compared consisted of the full address: the name of the street and the number of the building.

The arguments that the function takes are:

```
fuzzymerge(df1, df2, key1, key2, threshold=80, limit=1)
```

- df1: the left table to join
- df2: the right table to join
- key1: key column of the left table
- key2: key column of the right table
- threshold: how close the matches should be to return a match, based on Levenshtein distance
- limit: the amount of matches that will get returned, these are sorted high to low

The return of this function is a dataframe with both keys and matches. In order to later join the two datasets, it is necessary to merge them based on the *matching* column.

To be able to define the ideal threshold for the fuzzy joint, some test code runs were performed. This led to the selection of the first threshold equal to **88**. When the threshold was less, the algorithm was assigning street name matches too loosely, which was far away from reality, and while it was larger, it didn't recognize the match when the special characters or abbreviations occur.

Since the order of the arguments inside the function matters, the best approach in this particular situation was to fuzzy merge the cadastral data with the consumption data in order to prevent missing some crucial information about the electricity usage, and speed up the computation process.

During the stage of the fuzzy joint numerous issues occurred. Below is the list of problems and the way to overcome them.

1. **Building's Numbers**
   The fuzzy joint function counts how many characters match in between the con-

sidered addresses. For this reason if one digit of the number was the same in both datasets, it automatically considered it as a match.

```
CAMPDEVANOL ESTACIO 18 -> CAMPDEVANOL ESTACIO 1
```

In order to solve this problem, the address column had to be again separated into two: the street name and number of the real estate to be able to apply the fuzzy logic only on the street name, and after that, the numbers were compared outside the matching function. Since not always the exact number of the address appeared in both dataframes, a choice to take into account a small margin was made. The margin was set to 6, and in further steps, it could reach even 10. It means that the program would match the addresses with the same name and with the difference of the number less than 10:

```
CAMPDEVANOL ESTACIO 1 -> CAMPDEVANOL ESTACIO 5
```

This margin could be applied because of the 500 meters radius that has to be maintained within an energy community. It is the maximum distance, between a generation unit and a self-consuming point, that can be considered for an EC. Having this approximation in individual cases is then acceptable to be able to utilize the maximum amount of data that was provided by the municipalities. It was the trade-off that had to be taken, and for this specific case, it did not have a bad influence on the reliability and quality of later steps and finally the project itself.

Later on, the code only saved the cases with the smaller difference of numbers and the same street name to not have extra matches.

2. **Coordinates**
   Another problem that arose was that there were cases where one cadastral reference, meaning one building, had more than one set of coordinates. It suggests that the real estate was big enough to be assigned to a range of more than one coordinate set. It could be a problem while visualizing the final data with the GIS-based platform.

   Therefore, in order to solve the issue, the average value of the set of coordinates was taken into consideration, with the use of the following command:

```python
df.groupby(['ADDRESS', 'cadastralReference'])['xcoord',
    'ycoord'].mean()
```

3. **More than one cadastral reference**
   One of most significant issues developed throughout the process of the data harmonization was the case when more than one cadastral reference was assigned to the same address. In real terms, it indicates that one consumption value of one particular address was assigned to a broader number of the cadastral references,

and more than one sets of coordinates, like if it was more than one building. To make the final dataset as close to the truth and reliable as possible, the decision was made to take into consideration the surface area of each cadastral reference in order to proportionally split the consumption data. The automated process was developed using the Python script:

```
for index, row in filtereddf.iterrows():
    ratio=row['SUPTOTALCONSTRUIDAM2']/filteredgroupbydf.iloc[0]\
    ['SUPTOTALCONSTRUIDAM2']['sum']
    filtereddf.loc[index, 'JANUARY 2018'] = row['JANUARY 2018'] * ratio
```

It determined the surface area ratio for each legitimate case and then used that ratio to calculate the proportion of the power usage for the specific addresses. Additionally, to find the cases when this problem was occurring, the *dict.fromkeys*() method was used, which creates a dictionary from the given sequence of keys and values and gives out the output of each unique element within the specified column.

4. **Too broad match**
   A few cases appeared where in a cadastral dataset there was only data about one specific number on the street but it did not match the exact and also only available address in the consumption database. However, the fuzzy logic assign it as a match because of the high percentage of similarities, and no other options.

   `PALAFRUGELL FAR 2 -> PALAFRUGELL FAR 22`

   In the cases where the difference between two numbers was higher than 10, the match was deleted, similarly to the previous case with the Building's Numbers problem.

   Analogously, if there was only one position in the geo-location data, the code assigned all similar names from the electricity usage file:
   `PALAFRUGELL LLORET 14 -> PALAFRUGELL LLORET 2, 4, 6`

   To solve this problem, only the address which was the closest to the match was kept in the final dataframe, all others were deleted.

5. **Apostrophes & Special Characters**
   The code doesn't distinguish between different kinds of apostrophes. It had to be changed with the additional function outside the fuzzy logic. The same applied in the case of for example underscores.

```
df['STREET'] = df['STREET'].str.replace("'", "`")
```

All of these improvements were made, and the first run of the *fuzzy joint* was completed.

After the process' initial rounds, an additional **450** matching cases were extracted from the original dataset.

The important feature of the *fuzzy* function that makes it this attractive, is that it can be easily adjustable and changeable, which was utilized to improve its functionality and performance. In this particular thesis, it was to extract more hidden matches from the initial dataframe.

In further steps, the threshold was changed for each iteration in order to obtain as many matching pairs as possible. It was progressively decreased to make the matching process less strict. After applying the same technique as described above, the final **2148 matches out of 2409 of the original electricity consumption dataset, was found**. As it was previously explained, the addresses are not matching 1:1, although the similarity and close proximity make it still valuable for the context of this research.

The same exact reasoning was used in the instance of the **gas consumption**. The majority of the errors from the prior case are applied here as well. The *fuzzy logic* had to be performed a few times with different thresholds in order to extract every possible match. Starting from the original **1789** rows of the raw data, after applying merging functions, fuzzy joint, splitting the consumption, taking the mean value of sets of coordinates, and iterating through it several times, the final version successfully extracted **1523** matching cases.

6.2.4 Hourly Electrical Power Demand Curve

The **demand curve** is a visual representation of the variation in energy demand of consumers with regard to time. This curve is referred to as the daily demand curve if it is drawn over a period of 24 hours [48]. Most segmentation models for energy use rely on daily data. Given that the differences across load curves primarily occur on an hourly basis, this is a fair method. Different demographic and behavioral categories might be distinguished using the various signal patterns. At noon, the consumption proportion of elderly and parents of young children is larger than that of the typical consumer. Depending on the stage of the family life cycle, different nighttime power demand peaks may occur. Higher energy-awareness users could create smaller peaks and exhibit less fluctuation in their consumption share. If consumers pay attention to varied pricing and use more power during times when it is less expensive, energy tariffs in some countries may also have an impact on how much they consume [49].

In the context of this thesis it is important to extract this diagram in order to show the power demand of particular houses, to later compare it with the potential photovoltaic generation. After these operations it is possible to calculate the ratio of the demand and generated power, and withdraw important conclusions regarding the creation of a potential energy community. It is crucial to mention that different economic sectors characterize different typical power demand curves. Three main sectors can be distinguished: residential, commercial, and industrial shown on Figure 22.

The **residential** demand curve has a deep minimum at night when the major part of the

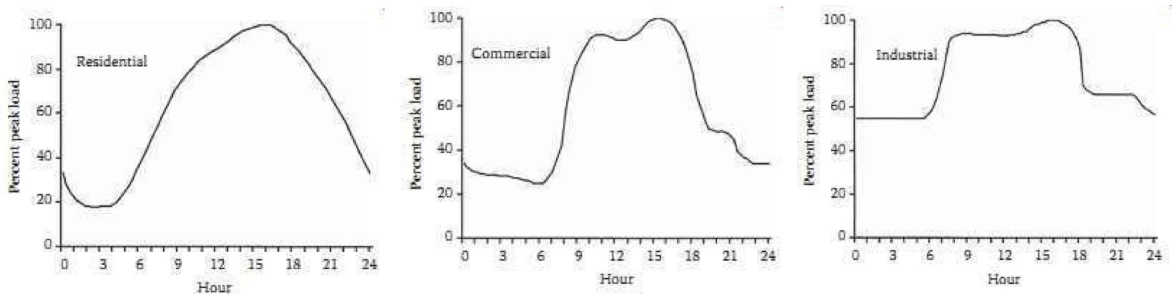

Figure 22: Typical Daily Power Demand Curve for Different Sectors

population is sleeping, so the consumption is very low. The load is distributed mainly from 7 a.m. to 10 p.m., and it has its peak coinciding with the typical end of the work time.

On the contrary, in the **industrial** curve the load is kept constant during the night at about 50% of the maximum load. Even though the main activity of the industries takes place during the day, there is an important part that still operates during the night, and approximately at 7 a.m., the consumption from the part of the industry only operating in diurnal time is added. In the industrial curve it is very clear that from 6 a.m. to 6 p.m., mainly working hours, the industry is operating at almost the maximum load in a constant form.

The **commercial** curve is similar to the industrial diagram, working hours can be recognized, but it has the characteristics between the residential and industrial curves. At night the consumption is very low, as a residential one, since the commercial sector does not operate during these hours. A small deep from 12 p.m. to 3 p.m. is observed coinciding with the pause of most shops, and also a small peak between 6 p.m. and 9 p.m. coinciding with the different times of closing of different types of shops and businesses.

In the case of this project, the hourly electricity consumption for the period of 3 years was downloaded from the *DataDis* platform. However, it was aggregated on a postal code level because of confidentiality. The original structure of the database is shown in the Figure 23 below:

postalCode,economicSector,tariff,dat	e,sumContracts,hour,energy,time	
17840,1_residential,2.0A,2018-05-01	00:00:00,1235.0,0,258.0,2018-04-30	22:00:00
17840,1_residential,2.0A,2018-05-02	00:00:00,1235.0,0,243.0,2018-05-01	22:00:00
17840,1_residential,2.0A,2018-05-03	00:00:00,1235.0,0,233.0,2018-05-02	22:00:00
17840,1_residential,2.0A,2018-05-04	00:00:00,1234.0,0,239.0,2018-05-03	22:00:00
17840,1_residential,2.0A,2018-05-05	00:00:00,1234.0,0,266.0,2018-05-04	22:00:00
17840,1_residential,2.0A,2018-05-06	00:00:00,1229.0,0,274.0,2018-05-05	22:00:00

Figure 23: Hourly Electricity Consumption Aggregated on a Postal Code Level

The dataframe included over 2,665,000 rows, it was a really heavy file that needed to be organized. As it is also visible, the Comma Separated Value file had difficulties to be opened correctly, which also was solved with the Python script by defining a delimiter. It consisted of the 8 parameters within which some other variables were included:

1. **Postal Code** - Codes of the only six municipalities included in the scope of the thesis
2. **Economic Sector** - 1_residential, 2_agriculture, 3_industrial, 4_1_office, 4_2_retail, 4_3_public spaces
3. **Tariff** - 2.0A, 2.0DHA, 2.0DHS, 2.1A, 2.1DHA, 3.0A
4. **Date** - 2019, 2020, 2021
5. **Sum of Contracts** - number of clients using the energy
6. **Hour**
7. **Energy** - energy consumed by the clients
8. **Time**

In order to make the computational process a bit faster and clearer, the database was split into smaller ones, based on the postal code. Additionally, each of the hourly data of different municipalities was divided taking into consideration different economic sectors. In this way, it makes the aggregation with the monthly electricity usage way easier and it allows the comparison between different sectors, which will be done later in this thesis. Moreover, the *time* parameter was converted into *datetime package* of Python.

Date and time are not separate data types in Python, but they may be used together by importing a *datatime package*. There is no requirement to install the Python Datatime Module externally because it is already included in Python. Datatime Module supplies classes to work with date and time. These classes provide a number of functions to work with date and time intervals. Python treats date and time as objects, not strings or timestamps [50].

Tariff aggregation
In hourly data, different tariffs were included for various numbers of contractors. It indicates that within a particular postal code area, and during specific hours, there were buildings that had different electricity tariffs than the rest of the buildings. Since the tariffs in the monthly electricity consumption dataset were not specified, all of the tariffs in the DataDis set had to be aggregated to one **universal tariff**, to be able to make it comparable. It was done separately for each of the economic sector cases so it could be differentiated easily. To do so, the *Panda's merging method* was used on the *time* column. The part of the code responsible for this operation is shown below.

```
import pandas as pd
```

```
df = pd.read_csv("17500Ripoll.csv")
df = df[df['economicSector'].str.contains("1_residential")==False]
df = df[df['economicSector'].str.contains("agriculture")==False]

df1 = df[df["tariff"].str.contains("2.0A") == True]
df2 = df[df["tariff"].str.contains("2.0DHA") == True]

dff = pd.merge(df1, df2, on = ["time"])

dff = dff.drop(columns = ['postalCode_y', 'date_y', 'hour_y', 'tariff_y'])
dff['tariff_x'] = dff['tariff_x'].str.replace('2.0A', 'UNI')

dff["sumContracts"] = dff["sumContracts_x"] + dff["sumContracts_y"]
dff['energy'] = dff["energy_x"] + dff['energy_y']
dff = dff.drop(columns = ["sumContracts_y", "energy_y", "sumContracts_x",
    "energy_x", "economicSector_y"])
dff.columns = ['postalCode','economicSector' ,'tariff', 'date', 'hour', 'time',
    'sumContracts', 'energy']

df = df[df['tariff'].str.contains('2.0A')==False]
df = df[df['tariff'].str.contains('2.0DHA')==False]

dff1 = pd.concat([dff, df])
dff1 = dff1.drop_duplicates(subset = ['time'], keep='first')
dff1["tariff"] = dff1["tariff"].str.replace("2.0A", "UNI")
```

In the first stage, all economic sectors and tariffs that are not of the thesis's interest were dropped with the *.str.contains* methods. Later, as mentioned before, the merging based on the *time* column was performed. The columns with the same values in both datasets (*postalCode, date, hour*) were dropped as well. In order to carry out the aggregation of the tariffs correctly, the *sumContracts* and *energy* values had to be summed up if it happened at the same time interval, then the duplicates were dropped. If during some hours there was only one tariff, it was just concatenated to the rest of the dataframe.

The logic was repeated for every tariff and for every economic sector, which made it possible to later compare with the monthly electricity data of particular buildings and extract daily load profiles for particular cases. The example of the dataframe with the tariff aggregation and cleaned data looks as follows in Figure 24.

Daily Power Demand

To extract daily or monthly electricity profiles for a particular address, two datasets had to be joined - *Monthly Electricity Consumption* & *Aggregated Hourly Electricity Consumption*. To do so, on the hourly data, the normalization by months had to be performed.

The first step to achieve this was the calculation of the electricity consumption of an average household, so the demand for a single building or real estate. It was computed

	postalCode	tariff	date	hour	time	sumContracts	energy
0	17200	UNI	2020-11-10 00:00:00	0	2020-11-09 23:00:00	10441.0	2310.0
1	17200	UNI	2020-11-13 00:00:00	0	2020-11-12 23:00:00	10441.0	2359.0
2	17200	UNI	2020-11-14 00:00:00	0	2020-11-13 23:00:00	10441.0	2506.0
3	17200	UNI	2020-11-15 00:00:00	0	2020-11-14 23:00:00	10441.0	2523.0
4	17200	UNI	2020-11-16 00:00:00	0	2020-11-15 23:00:00	10441.0	2304.0

Figure 24: Hourly Electricity Consumption with the Univeral Tariff

by dividing the *energy* by the *sumContracts*, and it was achieved by this line of code:

```
df['average'] = df['energy']/df['sumContracts']
```

Having this, it was possible to calculate the sum of the electricity consumption of an average household for each month in specific years. This value shows the accumulated consumption of a particular building in a specific month - on average this amount of energy was used by a sample building during a month period.

```
sum_average = df.groupby([df['time'].dt.year.rename('year'),
                          df['time'].dt.month_name().rename('month')])\
                          ['average'].sum().reset_index()
```

Later on, the *ratio* column was added which indicated the ratio of the hourly consumption and the monthly consumption. The value shows the fraction of electricity that was used at the specific hour in relation to the monthly usage. The sum of all the ratios is 1, it could also be expressed as a percentage for the simplicity. It was done with the following line of the script for each of the months.

```
may2018 = df.loc['2018-05']
may2018['ratio'] = may2018['average']/sum_average.average[5]
```

The final dataframe with the monthly normalization is shown on the Figure 25.

time	postalCode	tariff	date	hour	sumContracts	energy	average	ratio
2018-05-01 22:00:00	17200	UNI	2018-05-02 00:00:00	0	10339.0	2213.0	0.21404391140342394	0.0010872946829280335
2018-05-02 22:00:00	17200	UNI	2018-05-03 00:00:00	0	10338.0	2189.0	0.2117430837686206	0.001075606998670851
2018-05-03 22:00:00	17200	UNI	2018-05-04 00:00:00	0	10338.0	2273.0	0.21986844650802864	0.0011168820045586317
2018-05-04 22:00:00	17200	UNI	2018-05-05 00:00:00	0	10337.0	2419.0	0.23401373706104286	0.001188736882868048
2018-05-05 22:00:00	17200	UNI	2018-05-06 00:00:00	0	10337.0	2449.0	0.23691593305601238	0.0012034793824488835

Figure 25: Hourly Electricity Consumption Normalized by Month

The ratio defines the profile of the monthly consumption of buildings in a particular postal code area, in a specific economic sector, and with a *universal tariff*, and it is the same for every building within these specifications since the hourly data were aggregated on a postal code level. Hence, it is impossible to extract a specific demand curve for each building. It is an estimation of the load profiles supported and extracted from the real data provided by the municipalities. However, to make it more specific and relatable to each of the addresses, the fractions of the monthly consumption (*ratio* column) can be multiplied by the specific monthly electricity consumption of a particular address which will result in the same profile but scaled with respect to the real data of an individual real estate. Taking the electricity consumption of a specific month and year, and a specific address, and multiplying it with the previously calculated ratio, the *consumption* column can be added to the dataframe, which looks as follows in the Figure 26.

	time	postalCode	tariff	date	hour	sumContracts	energy	average	ratio	consumption
0	2020-01-01 00:00:00	17244	UNI	2020-01-01 00:00:00	0	4593.0	1662.0	0.361855	0.001071	2.462252
1	2020-01-01 01:00:00	17244	UNI	2020-01-01 00:00:00	1	4593.0	1511.0	0.328979	0.000974	2.238546
2	2020-01-01 02:00:00	17244	UNI	2020-01-01 00:00:00	2	4593.0	1374.0	0.299151	0.000886	2.035580
3	2020-01-01 03:00:00	17244	UNI	2020-01-01 00:00:00	3	4593.0	1276.0	0.277814	0.000822	1.890393
4	2020-01-01 04:00:00	17244	UNI	2020-01-01 00:00:00	4	4593.0	1226.0	0.266928	0.000790	1.816318

Figure 26: Sample Database of an Hourly Consumption of a Particular Address

In the above example the *ratio* column is multiplied by the monthly consumption of an individual building in Cassa de la Selva municipality which was equal to 2298.511 kWh. The result of each timestamp was saved in a *consumption* column.

At this point all the profiles can be extracted and plotted - the daily load curve and the monthly consumption profile. In the next section the profiles of different addresses, economic sectors and municipalities of different months will be presented, compared and analyzed.

6.2.5 Results

In this section the comparison of daily load curves during specific months will be presented separated by economic sectors. The graphs represent the Energy Consumption in kWh on the y-axis and the hour of a day on the x-axis, each of the lines in a figures is a visual representation of a electricity usage of one day.

Residential Sector
Below one can see the residential load curves for two distant months: January and August for one sample address in different municipalities. Each visible line represents the particular day in the month. In some figures there are visible some errors where one hour has a suspiciously low consumption. It is repeated in some cases because the shape of the profile is the same for the particular cases. It is important to mention that the curves are the approximation of an average household in a particular municipality, it is not a literal representation of a consumption of this specific building.

As it is shown in the Figure 27 below the magnitude of electricity consumption varies from case to case, and from month to month, but the profile shape is repeatable. The pattern that can be noticed it that the power usage is significantly higher during summer than during winter months. It is typical for the residential sector in hotter countries, which Spain is, because of the cooling systems which is always electric, in contrary to the heating systems, especially in Spain, which mainly is powered by oil, gas or diesel. The peak is around 10 a.m. when it starts getting hotter and people are still at home, then it decreases a bit when most of the residents are at work, and it goes back to the high consumption at 7 p.m. when everyone is coming back home. On the contrary, the winter shape of the curve is more flat, with the huge peak in the evening hours, around 7-8 p.m. It is caused by the fact that during the winter, the heating is more required in the evening, when people spend more time at home.

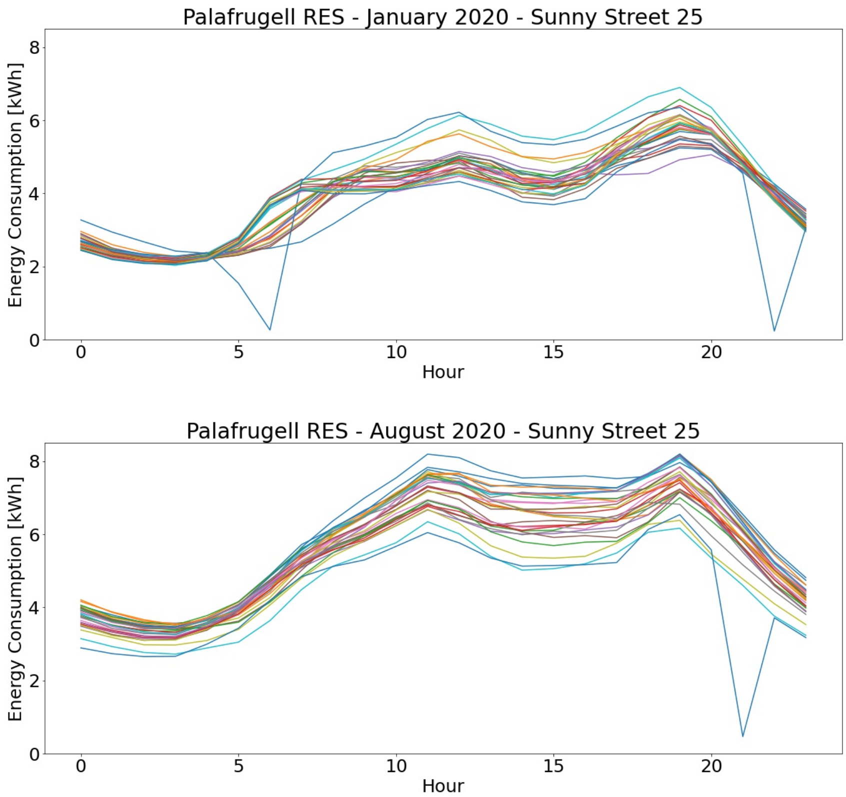

Figure 27: Daily Load Curves for the Residential Buildings - Palafrugell

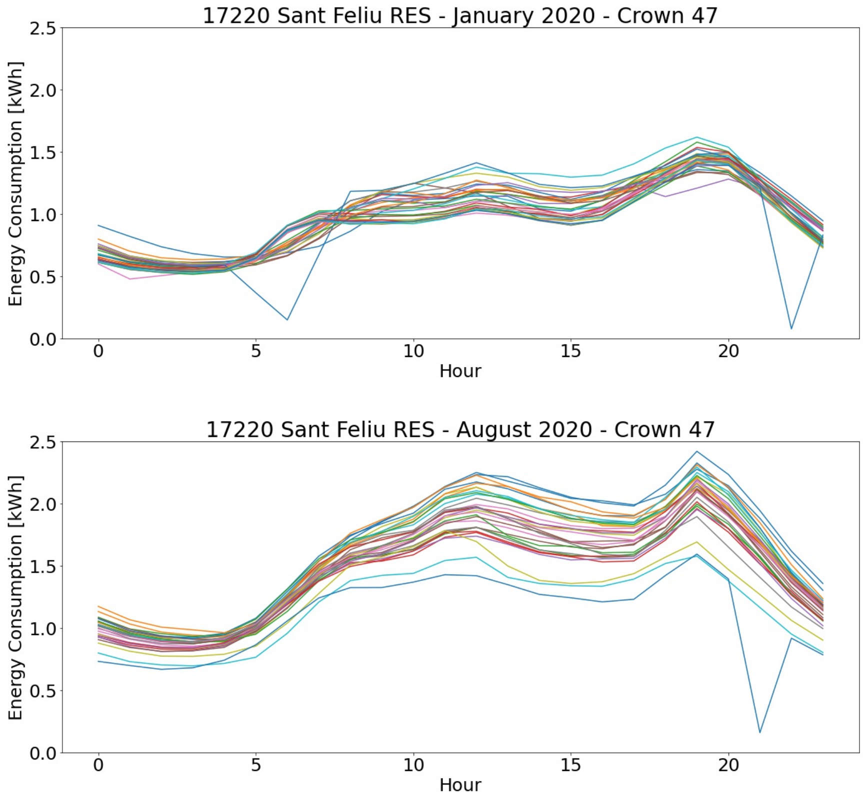

Figure 28: Daily Load Curves for the Residential Buildings - Sant Feliu de Guixols

Agriculture and Industrial Sectors
These two sectors were considered together in the same section because hourly data from DataDis did not differentiate them, in the contrary of monthly electricity consumption where it has been separated.

As it is noticeable, the shape differs a lot from the residential sector. Moreover, even both profiles from the same sectors are very different from each other, which shows that even within the same sectors there are many other factors that influence the behavior of the curves. People running the agriculture resort or a specific industry, area where it is located, a kind of resort among others have an impact on the profile.

In the case of the agriculture resort (Figure 29) it is strongly visible that the electricity consumption is only noticeable during peak hours, 8 a.m. to 8 p.m. During the rest of the hours it is either close to 0 or relatively low. In the case of the industries the consumption during the nights is not zero but it is kept constant at lower level. Additionally, the little deep in both cases in the afternoon around 1 p.m. to 3 p.m. is worth noticing. It is due to the small break in the middle of the day characteristic for Spain. Moreover, the weekends and holidays are very easy differentiable since the consumption during these days is also pretty low. When it comes to differences between January and August, it is other way around as in previous case in the industrial sector. The consumption is significantly higher during winter time than during summer (Figure 30).

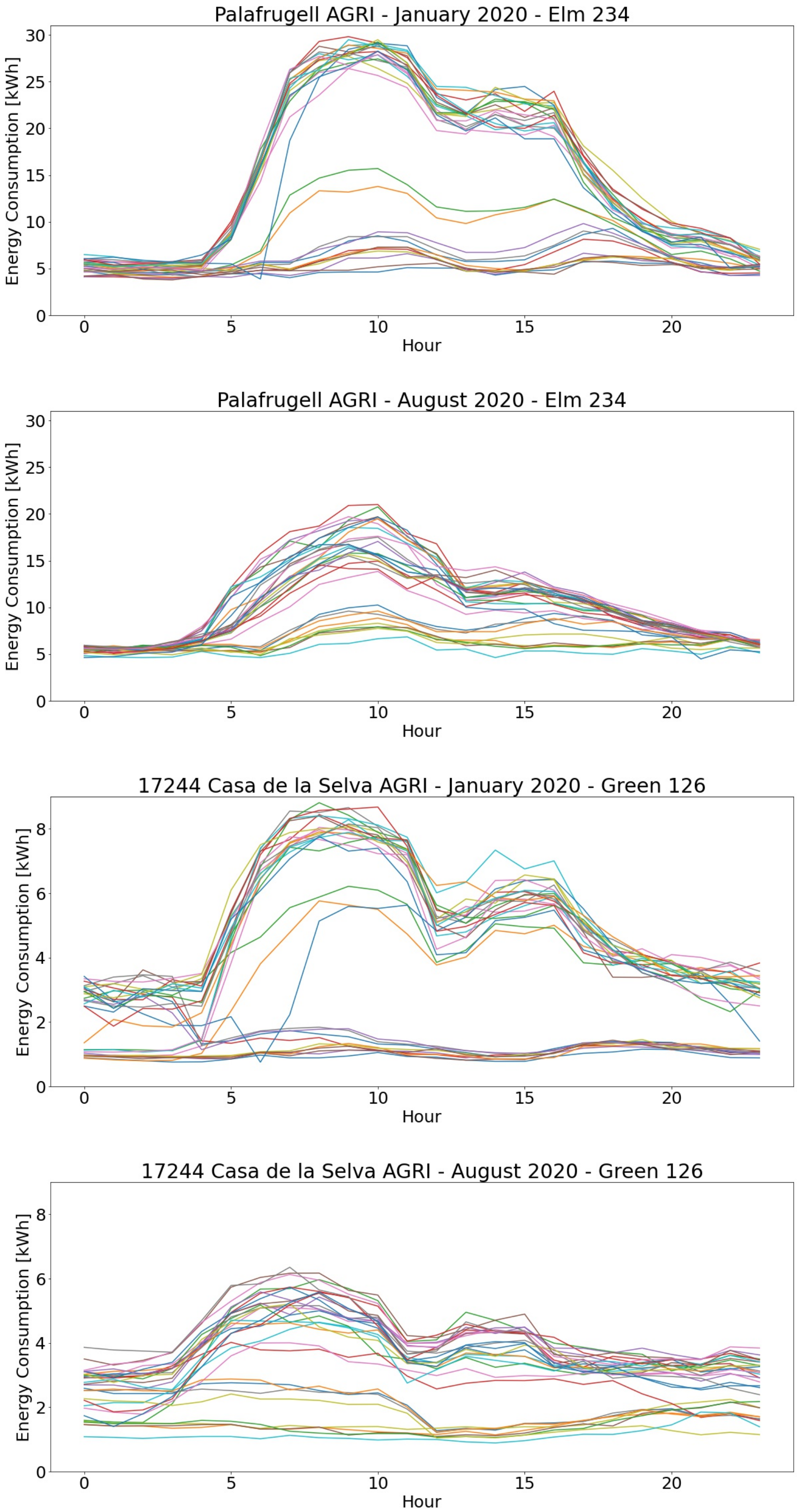

Figure 29: Daily Load Curves for the Agriculture Sector

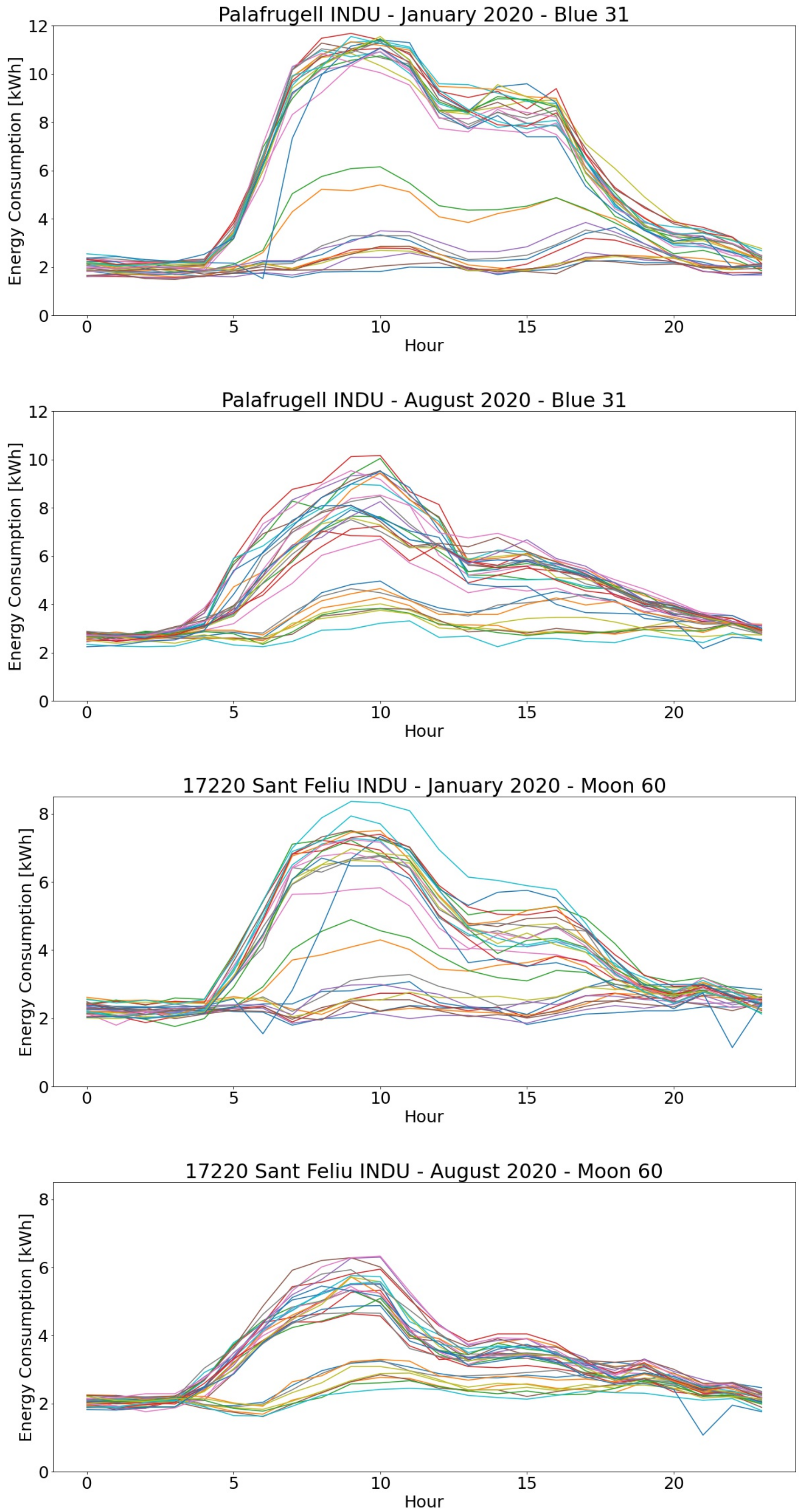

Figure 30: Daily Load Curves for the Industrial Sector

Office Sector

The daily load curve for the office sector (Figure 31 and Figure 32) resembles that for the industrial and agriculture sectors. It also has strongly highlighted working hours from 8 a.m. to 7 p.m. together with even more visible deep during lunch time, since the office sector is more influenced by this because if people don"t work, the energy usage is significantly decreased. Not like in the case of the industries, when even if people don't work, some processes still can take place. What is also similar to the previous case, is the fact that the weekends and holidays are very visible, because the energy consumption is significantly decrease and the shape is flat. This is because people don't come to the office during these days and they don't use electricity. Taking in to consideration the time of the year that the usage was measured, it depends on a particular case and the type of the office, similarly to the previous cases.

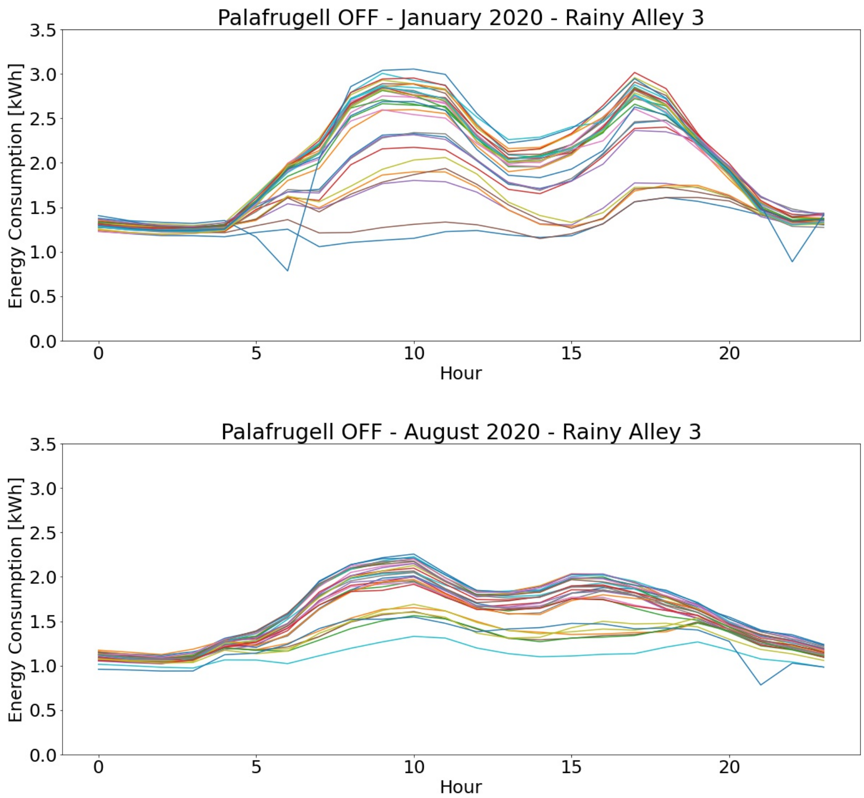

Figure 31: Daily Load Curves for the Office Buildings - Palafrugell

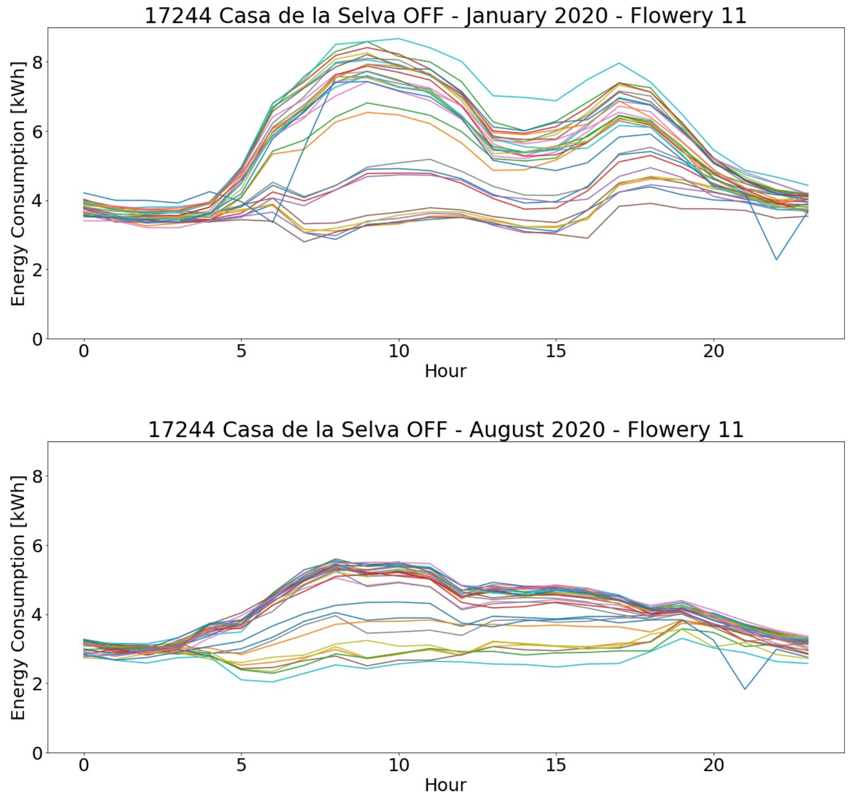

Figure 32: Daily Load Curves for the Office Buildings - Cassa de la Selva

As it is clearly visible in all the above cases, the electricity consumption profile varies from case to case, starting from economic sectors, through a time of the year, to the individual addresses. Although, the similarities can be noticed and a pattern can be drawn out.

Taking the profiles it is very easy to extract the behavior of the inhabitants when they wake up when they are using particular home appliances when they are gone etc. It is interesting that some *archetypes* can be highlighted from these consumers depending on their load curve pattern. On Figure 33 below some of them are shown. It shows how different the curve can be even when from the same sector.

These profiles can be approached efficiently in different ways, especially when taking into consideration the potential solar photovoltaic generation. Having different profiles causes better use of the self-produced electricity, therefore it is more profitable for the users than selling the electricity to the grid. It is even more amplified while considering the creation of an energy community. Since the electricity loss event is less probable. This is why it is important to include in an energy community consumers with different behavioral patterns that fit the specific generation. Mismatches are not desired because, on one hand, they can cause a situation when there is not enough electricity produced by renewable sources for all the clients - because the demand is higher than the production. In this case, the EC is not that beneficial because each member will consume

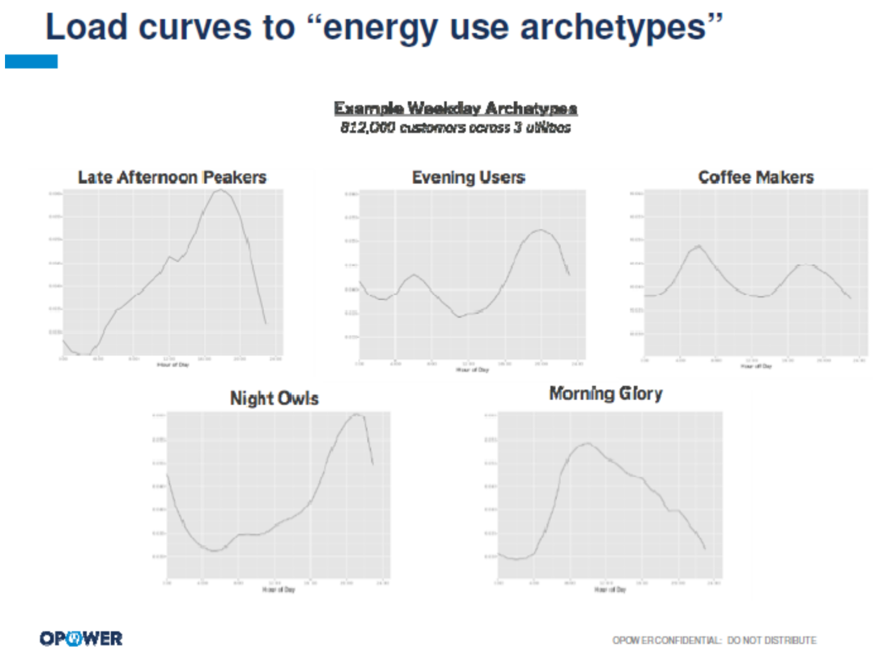

Figure 33: Load Curves to "Energy Use Archetypes"
[51]

only a small fraction of the generated electricity, while the rest would have to be bought from the grid. On the other hand, if the demand is high during hours the sun is not operating it can lead to some power loss.

One interesting phenomenon is called the **duck curve** which is shown on Figure 34.

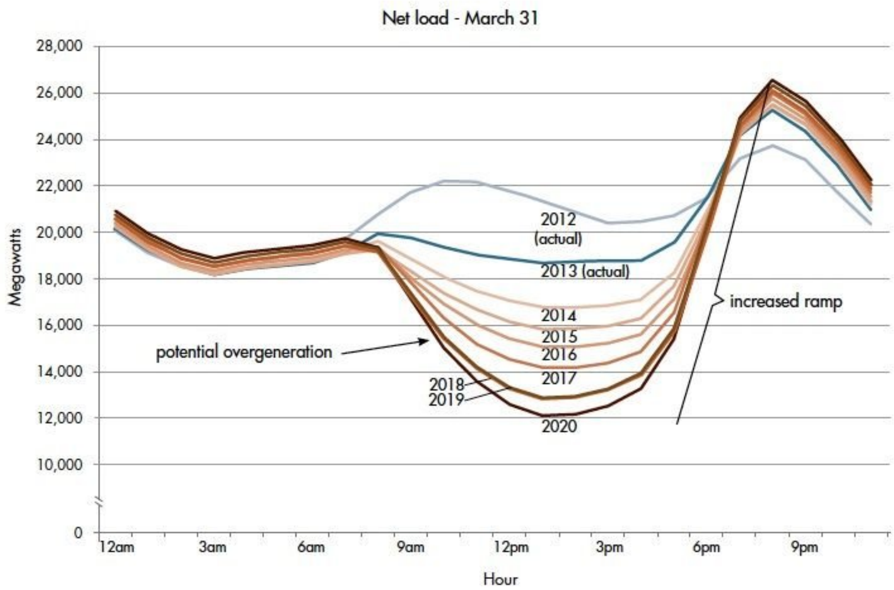

Figure 34: The Duck Curve
[52]

The gap between the daily power consumption and the quantity of solar energy available is represented by the duck curve, which got its name from how it resembles an animal. When the sun is shining, the solar dominated the market before declining as evening power demand increases. High solar penetration makes it difficult for utilities to maintain a balance between grid supply and demand. This is because as the sun sets and the output from PV declines, there is a greater demand for power producers to swiftly ramp up energy production. The potential for PV to create more energy than can be used at once or over-generation, is another issue with high solar adoption. As a result, PV production is reduced by system operators, diminishing its economic and environmental advantages. While curtailment does not significantly affect the advantages of PV when it happens periodically throughout the year, it may do so with higher levels of PV penetration.

The analysis of the results will be presented in the following section together with the conclusions withdrawal.

6.3 Results Interpretation and Discussion

Thanks to the full procedure of data collection, aggregation, and unification it was possible to extract daily load profiles of individual buildings from the pilot municipalities. One of the key features of this thesis is the fact that it gathers real data with different frequencies - hourly and monthly, and combines it with the geo-referenced data in order to be able to visualize it on the GIS-based platform. The project considered each of the economic sectors separately to be able to differentiate residential, industrial, and offices with public sectors and their characteristics. Unique characteristics define each of these sectors.

Daily load curves are very important when considering the creation of an energy community based in the solar PV system as a shared generation unit. It is due to the fact that both the demand curve and solar generation curve are structured and repeatable, hence they can't be adjusted. It is impossible to generate electricity through solar panels during the night nor to impose on people to use power only during 10 specific hours of the day. Figure 35 illustrates an example of a consumer's load profile, in the presence of solar PV daily production.

The orange is the PV generation and the blue one is the electricity consumption. The overlapping part is the PV power that is utilized directly within the building. This is sometimes referred to as the *absolute self-consumption*. The over-production is sold to the grid with the remuneration. It is important to understand the concept of that combined daily profile to be able to utilize the potential of an energy community to the maximum.

The idealized example of an Energy Community is when the whole PV generation is consumed by the members and the use of the grid is kept to the minimum. Hence, a variety of demand curves is required. As shown before in Figure 27, residential buildings are characterized by two peaks: during morning and evening hours, and a small deep in between them, which coincides with office hours, when people are usually not

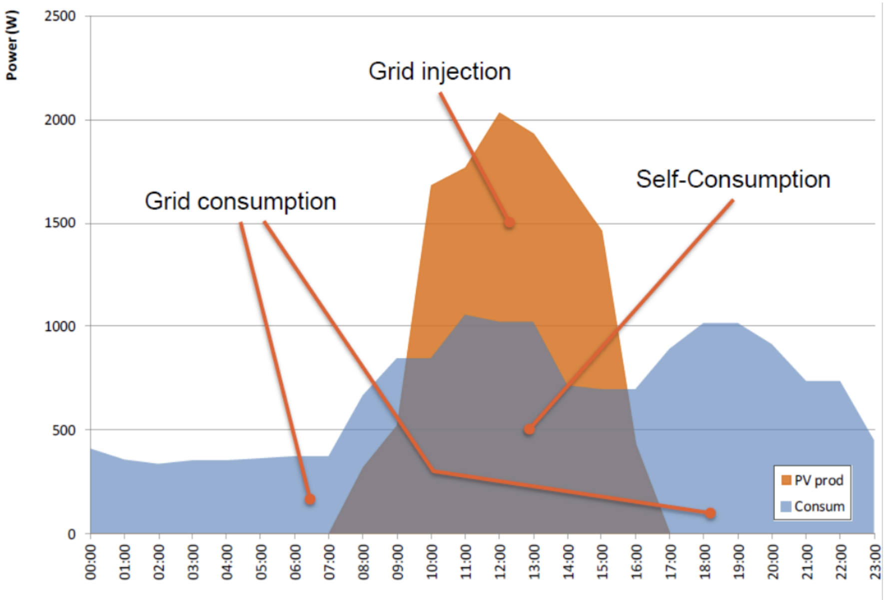

Figure 35: Typical prosumer's daily profile
[53]

at home. It is exactly the time when solar PV production is the highest. This is an event of the loss of opportunity because the excess power would be sold to the grid and not used by the members of the energy community. It would be remunerated but in the case of ECs it is way more beneficial for the users to self-consume the electricity rather than sell it to the system operator. Therefore, a good idea would be to include in this EC some industrial or office buildings which use the power constantly throughout the whole day and would compensate for the domestic consumers.

In that case, the question may arise, why not use only the industrial and agriculture sectors in the creation of energy communities? The demand profile of these sectors would be perfect, since during the night the consumption is very low, while during sun hours, the electricity usage is constantly high, without many deeps. It almost perfectly corresponds to the generation profile of the solar photovoltaic system. However, first of all, in most cases, there are not many industries or agriculture resorts within the range of 500m, because the residential sector is the majority within the cities. Second of all, and more importantly, as previously obtained plots (Figure 30) of industrial demand curves, the distinction between weekdays and weekends is very noticeable, because these sectors are closed during the weekends and holidays. Therefore, the generated electricity throughout these days would not be used and would have to be sold to the grid. It is enough reason to include diverse economic sectors in the creation of energy communities since similar motives can be easily found.

Another reason that confirms that energy communities should include within their members' various economic sectors is the fact that having the number of buildings of the same type would lead to the aggregation of the demand curves which have similar shapes. In this case, the aggregation would cause peaks would be even higher and deeps even lower, which could lead to the event of the previously mentioned duck curve (Figure 34) where the generation is way higher than the demand during afternoon hours. While during the time when the sun sets and it does not produce electricity anymore, the demand rises so high that it causes a very steep ramp which is very burdensome for the grid and system operators.

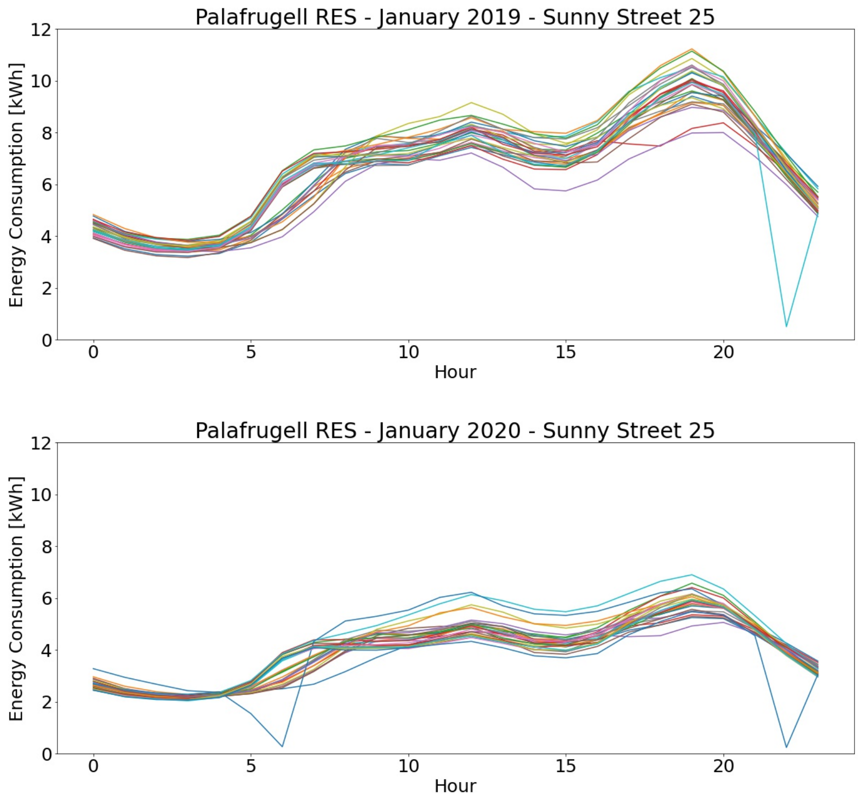

Figure 36: Daily load curves for different years - Palafrugell

As mentioned before the daily load profiles are static and vary over the time of the day as well as the month of the year. This aspect has to be taken into account while considering the creation of an energy community. Above on the Figure 36 and Figure 37 it is illustrated how the daily consumption profile changes for the same address, same month but in different years. It is crucial to investigate data from few years in order to forecast the power usage for the upcoming years. The purpose of this example is to show how dynamic the consumption can be, so even having data from many years does not mean that the forecast can be certain.

Nevertheless, it is possible for consumers to influence their behavior in order to utilize the PV solar production to the most beneficial level. Moreover, it is actually a PV installation itself and the concept of commitment to the EC that arouse the behavioral changes within the members. It is possible to speculate that a PV installation, either by itself or in conjunction with monitoring and visualization of electricity production and consumption, could arouse interest in the households' energy use and motivate efforts to further reduce it or to align it with the PV power generation, resulting in a difference

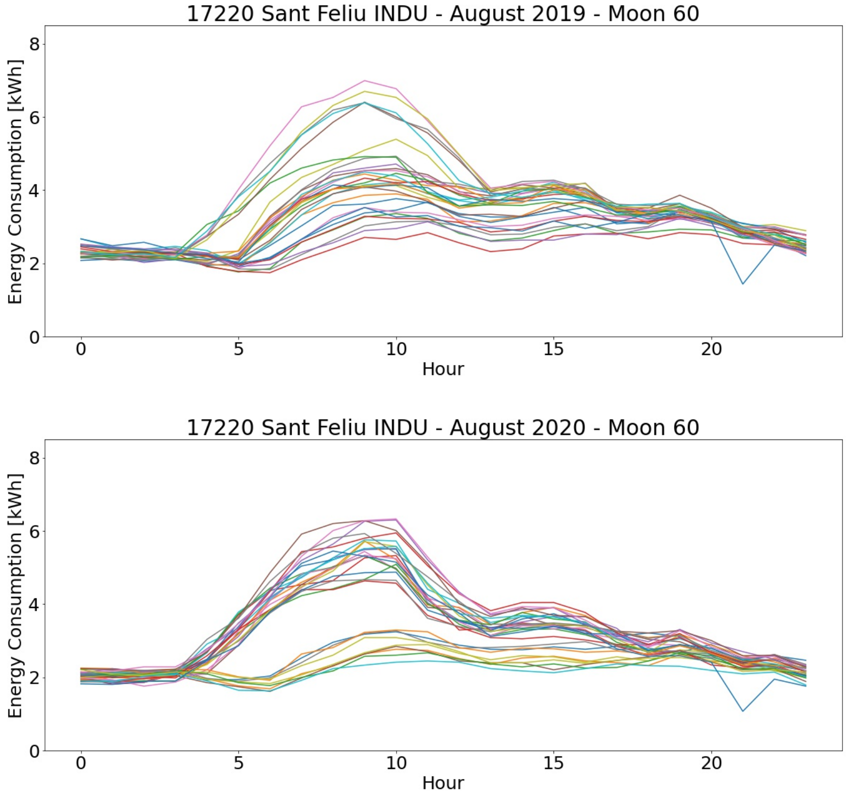

Figure 37: Daily load curves for different years - Sant Feliu de Guixols

between the total amount of electricity used and the daily load patterns before and after the installation. People seeing the amount of saved electricity, and specifically the saved money from not importing electricity from the grid, get more motivated to change their lifestyle patterns, Many studies prove this hypothesis right [54].

Not only the customers behavior can improve the self-consumption. Two main methods that positively influence the concept of self-consumption are:

1. Battery storage - better optimization of the utilization of the PV electricity because it can be used outside the sun hours, however it increases the initial costs a lot.

2. Demand Side Management - shifting of flexible electric loads in households to optimize the utilization of the PV production. Can be done with smart meters, load limiters or direct-load control.

Having listed many solutions for improving the self-consumption lead to the conclusion that even if the daily load profiles are not perfect and 100% compatible with the solar PV production, it still can be beneficial to create an energy community, because members feel more obliged and motivated to change their current patterns to match the PV generation more.

An energy community creation is beneficial for all its members for a longer run - it is unquestionable. It provides economic compensation since people don't have to buy that much electricity from the grid, and if there is an event of the excess energy, it is possible to sell it. It also gives the members a sense of belonging to the social circle and simultaneously helps fighting environmental crisis and helps achieving energy goals. Additionally, apart from financial, social and environmental benefits, energy communities can also help with an energy poverty problems by including vulnerable consumers in the community. Administrative works concerning this aspect are already ongoing.

However, energy community is still a relatively new concept. If it will get more popular it can cause some problems for system operators. The duck curve is the perfect example, where the demand rapidly increases when the sun hours are over because it is when people can't consume their self-produced electricity. The grid in such situation will be subject to huge fluctuations and tensions. This is not a problem directly related to energy communities but to increased number of renewable sources in general. The grid will definitely have to be reinforced or modernized in order to withstand these events. It will probably be related with the increased prices for the grid users. Automatically, users that are a part of the community will pay less than consumers that use the grid continuously and more often. This is the challenge which can occur in the further future and which will have to be faced by the grid operators and future legislation, but it is something that is worth keeping in mind - now is the perfect time for the creation of energy communities initiatives because the legislation has never been so favorable and susceptible to improvements.

6.3.1 Budget analysis

Since this thesis has developed as a part of bigger European project called ePLANET, which has 3 pilot areas: Girona, Zlin and Crete, it obtained a big financial support from the Europen Union in the about of € 1 498 533,75. The breakdown of this EU contribution between the project's participants is as follows:

1. Spain
   (a) Institut Catala d'Energia - €158 437,50
   (b) Diputacion de Gerona - €86 271,25
   (c) Associacio Lima - Low Impact Mediterranean Architecture - €168 916,25
2. Greece
   (a) Centre for Renewable Energy Sources and Saving Foundation - €145 750,00
   (b) Perifereiako Tameio Anaptyksis Kritis - €84 812,50
3. Czech Republic
   (a) Energeticka Agentura Zlinskhego Kraje Ops - €84 586,25

4. Belgium: Federation Europeenne des Agences et des Regions pour L'Energie et L'Environnment - €149 322,50

5. Germany: ICLEI European Sectretariat GMBH - €147 937,50

6. France: Three O'Clock - €152 500,00

The expenditures associated with this project's preparation include expenses for the employees who helped with its execution and for the software licenses required. The breakdown of the costs is shown below:

1. Staff: 1 person * 600 hours * €10/hour = €6000

2. Software - PyCharm: €90/year

The total cost to develop this thesis amount for €6090.

7 Conclusions, Contributions and Future Work

This thesis was written under the supervision of the independent department of CIMNE - BEE Group, which started developing a Horizon 2020 project called ePLANET in September 2021. The project will last for 3 years and will finish in August 2024. This work contributed to the very early stages of the mentioned H2020 project, mainly related to gathering data and data analysis, but also to the very initial phases regarding the general idea of the platform, state of the art, and discussions with potential collaborators and competitors. Each. of these stages is described in detail in the below section.

7.1 Main Contributions of the Thesis

As it was highlighted before the start of this work coincided with the start of the bigger European project called ePLANET which is part of the Horizon 2020 initiative. The early phases of the development of this thesis were getting a depth analysis of the concept of energy communities, as well as their regulation from the beginning - how it changed throughout the years and how it is now.

Having understood the concept of ECs better it was easier to conclude the direction of this thesis. At that point, the final idea about the ePLANET platform was not known yet. Hence, it was more than necessary to investigate the current stage of development of the sector. Meaning, analyzing already existing platforms related in any way to energy communities and conceiving an idea that stands out or contains more relatable features.

After analyzing 23 different platforms and companies which work to contribute to the development of energy communities, they were divided into two parts, which are the main points of interest for this thesis: these which are GIS-based, and the ones that are a tool for municipalities for the ECs creation. None of them included real consumption data with different frequency levels (monthly, hourly) together with the solar PV production potential of individual buildings. This allowed finalizing the clear idea of the whole ePLANET project, which will consist of 3 parts: an energy transition plan monitoring system, public energy transition measures database, and a GIS-based support tool for energy transition planning.

The latter is the main point of interest in this particular thesis. By establishing links with the appropriate data sources, the platform will perform harmonization operations and make it possible to visualize many layers of data. Additionally, users will be able to switch between other graphical scales, including building block, municipality, and postal code levels, and conduct benchmarking comparisons across areas of interest. To construct communal self-consumption projects for energy communities, one use of a GIS-based support system may be to connect energy consumption data at the building block level with PV rooftop potential studies. This thesis main contributions are regarding the data gathering, analysis, and harmonization of different types.

The subject of this thesis was data harmonization and visualization, together with the characterization of the buildings. All of it is a perfect base for future developments which will be explained in the next paragraph. 4 different datasets obtained from 6 different municipalities in Girona had to be unified, it included: monthly electricity consumption data, yearly gas consumption, aggregated hourly electricity consumption, and cadastral data. With the use of Python environment, Pandas library, and Fuzzy logic it was possible to obtain the daily load profiles of individual buildings. The developed script distinguishes economic sectors, from residential, through industrial and agriculture, to offices and public buildings.

Moreover, the extracted load profiles make it possible to compute the coefficients of the split of generated electricity between all its members, which is possible from the 1st of January 2022. These coefficients indicate the percentage distribution of energy produced from the solar PV system for each hour of the year - 8760 different values for one building. For that, the detailed data consumption of particular buildings is necessary to estimate the coefficients properly.

This developed tool is the base of the targeted creation of energy communities. With the generated energy profiles the size of the potential energy community and its components, as well as the distribution coefficients will be calculated and optimized.

7.2 Future Work

This thesis is a solid foundation for future work. Having daily load curves for each building based on real data, which is very innovative for the current state of the art, together with their geographical coordinates it is possible to visualize it on the map, which is one one the main points in the ePLANET platform.

Moreover, these load profiles can be crossed with the potential solar PV generation of particular rooftops, which in the case of this H2020 project might be done with a LiDAR technology to take into consideration the shapes, inclination, and shadings of the rooftops included in the pilot municipalities. Knowing the PV generation and the electricity usage it is possible to design collective self-consumption projects for energy communities.

It is a perfect tool for public authorities that can use public rooftops to install the PV system for the ECs, as well as an ideal incentive for citizens to see how much their own roof can produce energy and how much they would be able to save while becoming a member of an energy community. However, to optimize the size of the energy community it is also necessary to develop an optimization tool that would specify the size and the sharing coefficients within an energy community. This concept is already under development in the BEE Group, and it will use information from this thesis.

Bibliografia

[1] The Paris Agreement, https://unfccc.int/process-and-meetings/the-paris-agreement/the-paris-agreement, 2022

[2] Clean Energy for All Europeans, https://energy.ec.europa.eu/topics/energy-strategy/clean-energy-all-europeans-package_en, 2022

[3] Sustainable Development Goals, https://sdgs.un.org/goals, 2022

[4] World has installed 1TW of solar capacity, https://www.pv-magazine.com/2022/03/15/humans-have-installed-1-terawatt-of-solar-capacity/, 2022

[5] Luis Hernández-Callejo, Sara Gallardo-Saavedra, Víctor Alonso-Gómez, *A review of photovoltaic systems: Design, operation and maintenance*, August 2019

[6] Energy performance of buildings directive, https://energy.ec.europa.eu/topics/energy-efficiency/energy-efficient-buildings/energy-performance-buildings-directive_en, 2022

[7] Prof. L. Hancher and Mr. B.M. Winters, *The EU Winter Package. Briefing Paper*, February 2017

[8] Nearly zero-energy buildings, https://energy.ec.europa.eu/topics/energy-efficiency/energy-efficient-buildings/nearly-zero-energy-buildings_en, 2022

[9] Global Market Outlook for Solar Power 2022-2026, https://www.solarpowereurope.org/insights/market-outlooks/global-market-outlook-for-solar-power-2022, 2022

[10] Horizon 2020, https://research-and-innovation.ec.europa.eu/funding/funding-opportunities/funding-programmes-and-open-calls/horizon-2020_en, 2022

[11] ePLANET - European Public Local Authorities' Network for driving the Energy Transition, https://cordis.europa.eu/project/id/101032450, 2022

[12] ePLANET H2020 Official Website, https://www.eplaneth2020.eu/our-project/, 2022

[13] CIMNE - International Centre for Numerical Methods in Engineering, https://www.cimne.com/m2409/about/mission-and-history, 2022

[14] BeeGroup - Building Energy and Environment Group, https://www.beegroup-cimne.com/about-us/, 2022

[15] Energy Communities, https://energy.ec.europa.eu/topics/markets-and-

consumers/energy-communities_en, 2022

[16] Bettina Kampman, Jaco Blommerde, Maarten Afman, *The potential of energy citizens in the European Union*, September 2016

[17] Cristobal Gallego-Castillo, Miguel Heleno, Marta Victoria, *Self-consumption for energy communities in Spain: a regional analysis under the new legal framework*, January 2021

[18] Energy Communities Repository, https://energy-communities-repository.ec.europa.eu/energy-communities_en#map-of-energy-communities, 2022

[19] Aura Caramizaru, Andreas Uihlein, *Energy communities: an overview of energy and social innovation*, 2020

[20] Spain's energy cooperatives lead charge to exploit solar power, https://www.theguardian.com/world/2021/sep/01/spains-energy-cooperatives-lead-charge-to-exploit-solar-power, September 2021

[21] Energy Communities Repository, https://energy-communities-repository.ec.europa.eu/index_en, 2022

[22] Real Decreto 244/2019, https://www.boe.es/buscar/doc.php?id=BOE-A-2019-5089, 2019

[23] Javier López Prol, Karl W.Steininger, *Photovoltaic self-consumption is now profitable in Spain: Effects of the new regulation on prosumers' internal rate of return*, November 2020

[24] Som Comunitat Energetica, https://somcomunitatenergetica.cat, 2022

[25] Grid Singularity, https://gridsingularity.com, 2022

[26] Project Sunroof, https://sunroof.withgoogle.com, 2022

[27] AURORA - Smart Solar Power, https://aurorasolar.com, 2022

[28] Roofpedia, https://ual.sg/project/roofpedia/, 2022

[29] Tetraeder.solar, https://solar.tetraeder.com/de_v2/, 2022

[30] KM0 Energy, https://km0.energy/en/, 2022

[31] Power Quartier, https://www.exnaton.com, 2022

[32] SimStadt, https://www.hft-stuttgart.de/forschung/projekte/abgeschlossen/simstadt, 2022

[33] EnerMaps, https://enermaps.eu/news/enermaps-project-a-new-open-energy-data-tool-to-accelerate-the-energy-transition/, 2022

[34] Statistical Institute of Catalonia, https://www.idescat.cat/emex/?lang=en&id=170792; 2022

[35] Covenant of Mayors - Signatories of Covenant of Mayors, https://www.covenantofmayors.eu/about/covenant-community/signatories/overview.html?scity_id=12339; 2021

[36] Covenant of Mayors - Covenant Initiative, https://www.covenantofmayors.eu/about/covenant-initiative/origins-and-development.html; 2022

[37] Covenant of Mayors - Covenant Community, https://www.covenantofmayors.eu/about/covenant-community/signatories.html, 2022

[38] Covenant of Mayors - Province of Girona, https://eumayors.eu/news-and-events/news/1132-province-of-girona-over-90-of-municipalities-supported-in-reaching-covenant-climate-and-energy-objectives.html, 2013

[39] ManagEnergy - Building Sustainable Energy Investments in Girona, https://eumayors.eu/news-and-events/news/1132-province-of-girona-over-90-of-municipalities-supported-in-reaching-covenant-climate-and-energy-objectives.html, 2018

[40] Intituto de Estadistica de Cataluna, https://www.idescat.cat/emex/?id=170792&lang=es, 2022

[41] Python Documentation - General Python FAQ, https://docs.python.org/3/faq/general.html#what-is-python, 2022

[42] Libraries in Python, https://www.geeksforgeeks.org/libraries-in-python/, 2021

[43] ENDESA, https://www.endesa.com, 2022

[44] Nedgia, https://www.nedgia.es, 2022

[45] DataDis, https://datadis.es, 2022

[46] Sede Electronica del Catastro, https://www1.sedecatastro.gob.es/CYCBienInmueble/OVCConCiud.aspx?UrbRus=U&RefC=4790502DG8449B0015HS&RCCompleta=&via=ANTONI@VARES@MARTINELL&tipoVia=CL&numero=2&kilometro=&bloque=&escalera=&planta=1&puerta=5&DescProv=GIRONA&prov=17&muni=85&DescMuni=GIRONA&TipUR=U&codvia=1176, 2022

[47] A Simple Guide to Data Preprocessing in Machine Learning, https://www.v7labs.com/blog/data-preprocessing-guide, 2022

[48] A Load Curve, https://www.electrical4u.com/load-curve-load-duration-curve-daily-load-curve/, 2020

[49] European Commission, *Load Profile Classification - European Commission*, 2016

[50] Python Datetime Module, https://www.geeksforgeeks.org/python-datetime-module/, 2021

[51] European Commission, Horizon 2020, *D4.1 Load Profile Classification - WP4 – Classification of EU residential energy consumers*, July 2016

[52] Confronting the Duck Curve, https://www.energy.gov/eere/articles/confronting-duck-curve-how-address-over-generation-solar-energy, 2017

[53] European Photovoltaic Industry Association https://www.solarpowereurope.org

[54] A. Jenny, J.R.D. López, H.-J. Mosler, *Household energy use patterns and social organisation for optimal energy management in a multi-user solar energy system*, 2006