UNIVERSITY

SCHOOL OF SCIENCE & TECHNOLOGY

# CAPSTONE PROJECT

# ELECTRITECT

Master in Big Data and Business Analytics

School of Science & Technology

Academic Year 2023-24

*Key words:*

Energy optimization, energy consumption, cost reduction, forecasting

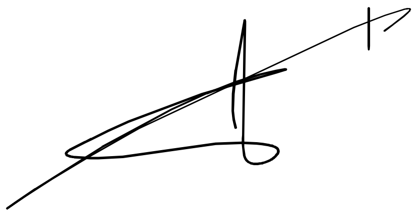
*Thesis Supervisor's Name:* Dr. David Yeregui Marcos del Blanco

Mikel Albisu Astigarraga:

“I hereby certify that this report and the accompanying presentation is my own original work inits entirety, unless where indicated and referenced.”

Andreu Artigues Martínez:

“I hereby certify that this report and the accompanying presentation is my own original work inits entirety, unless where indicated and referenced.”



Gabriel Briones Loría:

“I hereby certify that this report and the accompanying presentation is my own original work inits entirety, unless where indicated and referenced.”

Johan Barreiro

“I hereby certify that this report and the accompanying presentation is my own original work inits entirety, unless where indicated and referenced.”

# Contents

# 1 Introduction

Electricity has become a universally indispensable form of energy, essential for diverse applications spanning residential, commercial, and industrial sectors globally. Its role extends from enhancing quality of life and economic development to powering essential services and technological advancements in both developed and developing nations [1]. It is nearly impossible to replace its uses and applications; therefore, it can be said that well-developed societies heavily depend on the availability of electric energy, which has become an essential commodity [2].

As global economies expand and technology progresses, the demand for electricity continues to soar, particularly in large corporations, many of which form the backbone of industrial activities around the world. These entities are major consumers of energy, driven by their extensive operational scopes and the need to maintain continuous production cycles. Any surge in energy demand not only strains limited resources but also elevates their operational costs significantly, affecting the overall economic stability of these corporations [3].

## 1.1 Current Situation for Large Corporations

The regulatory landscape around energy consumption is becoming increasingly stringent, as governments worldwide implement policies to promote energy efficiency and reduce greenhouse gas emissions [4].Thus, large corporations find themselves at a crossroads; they must comply with these regulations while also seeking ways to optimize energy use and reduce costs. Failure to adapt to these regulations not only risks financial penalties but also reputational damage. This emphasizes the need for sophisticated energy management systems that can navigate these complex challenges.

These organizations, characterized by their extensive operational facilities and high energy demands, are major contributors to global energy consumption, accounting for nearly 60% of total industrial energy usage [5]. This substantial energy demand not only increases their operational costs but also significantly impacts their environmental footprint.

In addition, studies indicate that energy expenses can constitute up to 30% of the total operational costs in energy-intensive industries such as manufacturing and data centers [6].

From a regulatory perspective, the situation is equally demanding. The introduction of stringent environmental regulations, such as the EU's Green Deal, mandates a 32.5% improvement in energy efficiency by 2030 for industrial sectors, pushing corporations to adopt more sustainable and efficient energy practices [7]. Compliance with these regulations is not merely a legal obligation but a critical component of corporate social responsibility strategies, that influence public perception and overall investor confidence.

## 1.2 Our Sector: Industry 4.0

In recent years, industrial companies have increasingly grappled with the challenge of managing high energy consumption, which accounts for roughly half of all global energy usage. Their significant energy demand, coupled with fluctuating costs and stringent environmental regulations, has highlighted the need for optimization in energy use [8]. Industrial operations frequently encounter inefficiencies, as such, a considerable amount of energy wasted is through outdated systems and processes. These factors underscore the urgent need for robust energy management strategies in the industrial sector, paving the way for the Industry 4.0 [9].

This term refers to industries whose objective is to increase productivity and enhance mass production using innovative technology as well as advanced planning and control methods, leveraging a broad span of state-of-the art technologies based on Big Data, Cloud Computing and Smart Sensors, among others.

Smart Sensors, also referred to Internet of Things (IoT) technology, is a formidable tool for industries to improve monitoring, control, customisation, and prediction [10]. Specific to energy management, smart meters primarily feature data storage and data management capabilities as well as end-to-end communication and wireless connections [11, 12]. These features enable smart meters to become an important data source when monitoring energy consumption, alongside statistical prediction methods and machine learning. Given the current context that industries face, the current team developed a solution designed to enhance market value.

## 2 Electritect

In order to overcome the problems stated, our team came up with **Electritect**. Electritect aims to be an innovative Software as a Service (SaaS) designed to optimize energy consumption and enhance energy performance without affecting production levels or service quality. By identifying inefficiencies and suggesting targeted improvements, Electritect target medium and big corporations to help them:

- Reduce Operating Costs: By optimizing energy consumption, businesses can significantly reduce their operating costs associated with energy usage. This can lead to improved profitability and increased competitiveness. In addition, we can help them identify and eliminate inefficiencies in their energy usage. This can lead to improved productivity and reduce downtime.
- Enhanced Environmental Sustainability: Optimizing energy consumption helps businesses reduce their carbon footprint and contribute to a more sustainable future.
- Compliance with Regulations: Many businesses are subject to regulations that limit their energy consumption. Energy consumption optimization algorithms can help businesses comply with these regulations and avoid penalties as stated in subsection 1.1.

As mentioned, the core of our solution lies in the optimization of the energy usage. The optimization algorithms are designed to find minimum values of mathematical functions by selecting the best elements from some set of available alternatives. This approach will not only minimize power consumption but also reduce the final price cost.

To achieve this, implementation of hardware equipment (IoT) is needed. Since global energy readings alone do not provide significant advantages to the information that energy companies possess, it is necessary to go further and acquire more information about the infrastructure. By integrating with existing industry facilities, we can gain access to valuable data sources that enhance the effectiveness and performance of our solutions.

The more information we can gather and analyze, the better insights we can provide to our clients, empowering them to make data-driven decisions and drive business growth.

# 3 Proof of Concept

In order to validate our business model, we are using this Capstone Project to develop the technical aspects for Electritect. To accomplish this, we conducted a research study to identify a dataset that could effectively showcase the potential of our ideas. However, the dataset need to meet a specific set of requirements.

- The dataset should belong to a medium to large-scale infrastructure to align with our scalable solution for our target market.
- The dataset should contain detailed operational data from buildings, including electricity consumption measured by smart sensors across different sectors within the industry.
- Additionally, data related to indoor environmental measurements like natural gas consumption or water usage, collected from IoT sensors, would be beneficial and taken into consideration,

After performing an extensive search, we decided that we could actually perform our study on the following Industrial Site Dataset, containing real-world data related to energy consumption from a three-site industry from the European Union.

## 3.1 Industrial Dataset Description

We are considering this dataset for several reasons, in addition to those mentioned earlier:

1. During Venture Lab sessions, it was recommended that we explore sub-sectors within the same industry to demonstrate scalability in our business solution.
2. This dataset belongs to a market context closer to the team's target size. Since this data belongs to EU Industry, it is reasonable to assume that the data already is aligned or is on the process of aligning with EU criteria and regulations.

Furthermore, the election of this specific dataset is because most of the other datasets analysed only provide total readings for households or buildings but did not include inner consumption values[1]. Therefore, this Industrial dataset freely available online, best meets the specified criteria.

Lastly, this dataset has been developed by Energenius Srl, which is an innovative start-up specializing monitoring and analysis of energy metrics. This dataset was made public on October $4^{th}$ 2023, making it relatively new, which further supports the selection of this dataset. Unfortunately, due to privacy considerations, specific details about the industry or client to which this data belongs are not disclosed.

[1] Other additional datasets were discussed but ended up being discarded. Refer to Appendix A to see these other alternatives.

# 4 Optimization Algorithm

In this section, it is discussed the core of the business solution: optimization. This process consists of several essential components, each playing a critical role in minimizing the energy cost of the industrial site. The key elements include the Objective Function, Constraints, Bounds, and Parameters. Together, these components fit together like pieces of a puzzle, forming a cohesive and efficient optimization model.

Given the complexity, especially in large-scale industrial applications, standard Python libraries may not be sufficient for conducting advanced optimization. For this project, Electritect has chosen to use **Pyomo**, a Python-based open-source optimization modelling language. Pyomo supports both linear and nonlinear programming, making it well-suited for handling the intricate optimization tasks required in our project. Let us proceed to understand the different steps to achieve optimization.

## 4.1 Objective Function

The objective function is the cornerstone of our optimization model, representing the goal to be achieved and how other features relate to it. One could say that the cost associated to energy corresponds to the following equation:

$$Cost = \sum_{h} \sum_{i}^{N} P_h \times E_{i,h} \tag{1}$$

where $P_h$ is the set price of hourly energy and $E_{i,h}$ represents the energy consumption of each individual variable/segment of the industry, as measured by smart sensors, also per hour. Each segment could be a machine, a production line, or any other subunit that consumes energy. So, by aggregating through a certain period of time, one would obtain the total final costs.

## 4.2 Decision Variables & Parameters

Next, we have the Decision Variables, which are actually under our control. For example, we cannot dictate how much an air conditioner consumes[2] but we can control how long it can run and at what temperature.

On the other hand, there are the Parameters. As opposed to decision variables, we do not have any control over parameters neither by choice nor by nature. These variables set the conditions for optimization to obtain differing results. In Table 1, there are some examples to what both terms refer to.

## 4.3 Constraint Equations

Constraints are the crucial elements that link decision variables and parameters to the objective function, shaping the feasible solution space for the optimization model. These constraints ensure that the solutions generated are not only optimal but also practical

[2]Naturally, investment in latest pieces of technology that ensure minor consumption is a strategy at hand, but for this project and facing probable real life scenarios, we will consider only the equipment at hand.

| Decision Variables | Parameters |
|---|---|
| Temperature Set Point | Compressors Energy Consumption |
| Number of Active Chillers | Technological Centers Energy Consumption |
| Number of Active Coffee Machines | Energy Prices per MWh |
| Active Printers | Production Schedule |
| Energy for Transport Vehicles | Maintenance Status |
| Machinery Activity Status | Volume Production Waste |
| Number of Active Servers | Number of Workers |
| Number of PoE Network Switches | Chiller Power Factor |
| Number of Non-PoE Network Switches | Operational Presence |
| Number of active HDD | Fabric in Chamber |
| Number of active SSDs | Testing Schedule |
| Operational Presence | Workload |
| - | Heat Index |
| - | Active Wall Plugs |
| - | Active Light Bulbs |
| - | Active Wall Computers |

Table 1: Examples regarding different decision variables and parameters that are expected within industry contexts.

and adhere to real-world limitations. This is where the data comes in. Constraints often will come in the form of linear equations, although one of the reasons Pyomo is used is to have the ability for boolean operators in these equations.

Data was used to fit equations to solve for variables of interest. Here, the equations are the model's variables found in the objective function. One could say that any constraint equation, which aims to represent the sector/variable ($E_i$) consuming energy, will be the linear combination of both the decision variables ($DV$) and the parameters ($Params$) with a solution greater or equal than 0.

$$E_i = a_i + \sum_{k}^{K} b_k DV_k + \sum_{j}^{J} c_j Params_j; \quad E_i \geq 0 \tag{2}$$

## 4.4 Bounds

Bounds play into decision variables, constraints, and others. They dictate how the math needs to behave to simulate real world scenarios. Otherwise, the solver would simply push everything to 0 or lower.

Some bounds that are common to use along others are:

- Set Point may not fall outside 18 ºC and 23 ºC. If the heat index is 18 or under it will be set to 18. If the heat index is above 40 it will be set to 25. Otherwise, it will increase linearly between 18 and 25.
- Hardware in the Data Center (servers, network switches, hard drives) must be active in great quantities during peak hours.
- Energy consumption cannot be below zero.

- Office equipment (printers, coffee machines) must be fully active during office hours.
- Transport Vehicles charging and Machinery Activity status depend on energy price and time of day.
- Number of active chillers must be above 0 to maintain air flow.

### 4.5 Forecasting

To achieve future savings with historic data, forecasting techniques for parameters can be utilized. For some specific values, like energy price or weather data, these can be obtained from differing sources. OMIE, which is the Spanish entity in charged of the electric pool market, publishes hourly day ahead energy prices. Then, Visual Crossing API can be called to obtain tomorrow's weather data.

However, other aspects might not be able to be predicted that easily, for which several forecasting options will be considered:

- ARIMA(5,1,0)
- ARIMA(8,1,0)
- SARIMA(p,d,q)(P,Q,D)$_s$
- Random Forest Regressor transformed with sinusoidal feature.

In the end ARIMA(5,1,0) was selected as they better modeled the actual data.

So, once all components of an optimisation problem are set, it will be ready to be solved. Consequently, an analysis of the dataset will be presented, accompanied by an in-depth analysis of their features.

## 5 Exploratory Data Analysis

As with any data-driven project, it is essential to first fully understand the given data, previously presented in section 3. First, all three datasets contain time series information regarding different variables with a time granularity of 15 minutes. The overall length of the datasets include 42,760 records spanning from July $1^{st}$ 2022 at 00:00 to July $19^{th}$ 2023 at 09:45. However, it turns out that in all three files, the data collection stops on June $30^{th}$ 2023 at 23:45, resulting in a final dataset of 1-year timeline. Although the length of this time series is more than what we actually need, it is pertinent to keep it since this offers more information on aspects like seasonality, vacations periods, maintenance shutdown and periods of high demand.

### 5.1 Variable identification

All three dataset contained information regarding the monitoring of the sensors for different variables. The complete list of unique variables and their units are the following: Energy (kWh), Flow Rate ($m^3/h$), Power (kW), Pressure (bar), Power factor (real), Temperature (°C), Efficiency (%) and Quantity ($m^3$). Moreover, each site contained between 27 and 52 variables[3].

[3]The complete list of variables can be seen in Appendix C

Despite the vast number of variables present, the team identified the variables for each site that tend to repeat in format. It can be elaborated a format that will help quickly identify the general structure behind each dataset:

- First, a certain number of sectors unique to that dataset are shown using different measurement systems which are: active energy (kWh), power factor and active power (kW).
- Then, each dataset will present unique "global measurements", like `General Electric Active Energy, General Natural Gas Quantity, General Vapour Fumes Temperature, General Water Flow Rate and General Vapour Pressure` among others.

## 5.2 Datasets Inspection

Initially, it was conducted an exploration of missing values in the dataset. Surprisingly, Industry Site 2 presented a distribution of missing values of around 10% of the length of the dataset. On the other hand, Industry Site 1 & 3 only presented in very few variables missing values only accounting at maximum 0.3% of the dataset.

Furthermore, it was necessary to address outliers present in the dataset. While exploring the dataset, some of the outliers or inexplicably large values present only seemed to be punctual error signals, with enormous values. To manage these, values greater than the 99th quantile range were replaced with the 'forward fill' method by replacing these with the previous value. This also served to fill null records.

## 5.3 Relationships Encountered

After careful examination of the data, we realized that the dataset's structure suggested potential correlations between variables. As a result, we present the most relevant findings we encountered.

### 5.3.1 Monitored vs Theoretical Energy

As previously mentioned, all datasets contained information regarding a certain number of variables but in different measurement systems: active energy (kWh), power factor and active power (kW). These units indicate the following.

- Active energy: Total amount of electrical energy consumed over a period of time. It is measured in kilowatt-hours (kWh).
- Active Power: Instantaneous measure of power consumption. It is measured in kilowatts (kW).
- Power factor: the ratio of real energy usage / theoretical capabilities. It is expressed as a ratio or (%).

Due to the formal definition of these, one might expect that these variables are related to each other. If data is recorded every 15 minutes, which is 1/4hour or 0.25 hours, then:

$$Energy = Power \times 0.25 \tag{3}$$

And if on top of that, we know that energy efficiency is taken into account, the expected result should be modified by the power factor.

$$Energy = Power \times 0.25 \times PowerFactor \tag{4}$$

However, this theoretical formulation did not seem to be consistent, as it can be seen in examples Figure 1 & Figure 2. The election of sectors was not intended, as it was part of the data analysis. While in the first one values are close together, they are not exactly the same and the difference is notable[4]. In the second one, there is a clear discrepancy for some variables.

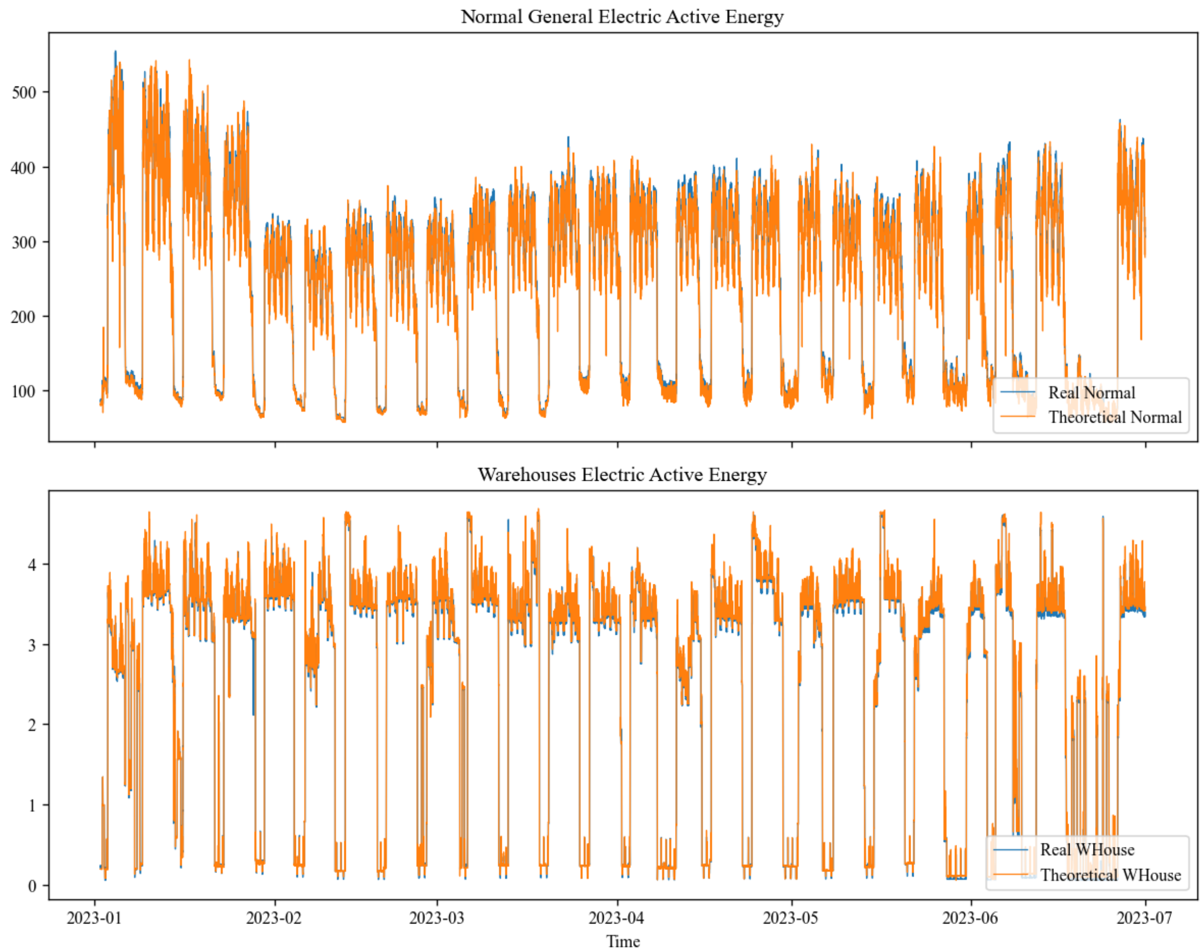



Figure 1: Differences between the Active energy obtained from the smart sensors vs the predicted ones in two sectors for Industry Site 1 for the last 6-month data.

Out of this figure, the thoughts are: Either the assumption in Equation 4 is incorrect and doesn't apply, or the data collection from the smart sensors does not perform in a precise and expected way. Regarding the first point, it might not be the case since we have proved that some sectors seem to adjust to the measured value. However, this observation does not discount the possibility that the data collection and the smart sensor readings might introduce errors.

[4]Later is was seen that within the same Industry Site some variables presented clear discrepancies, just as in Figure 2.

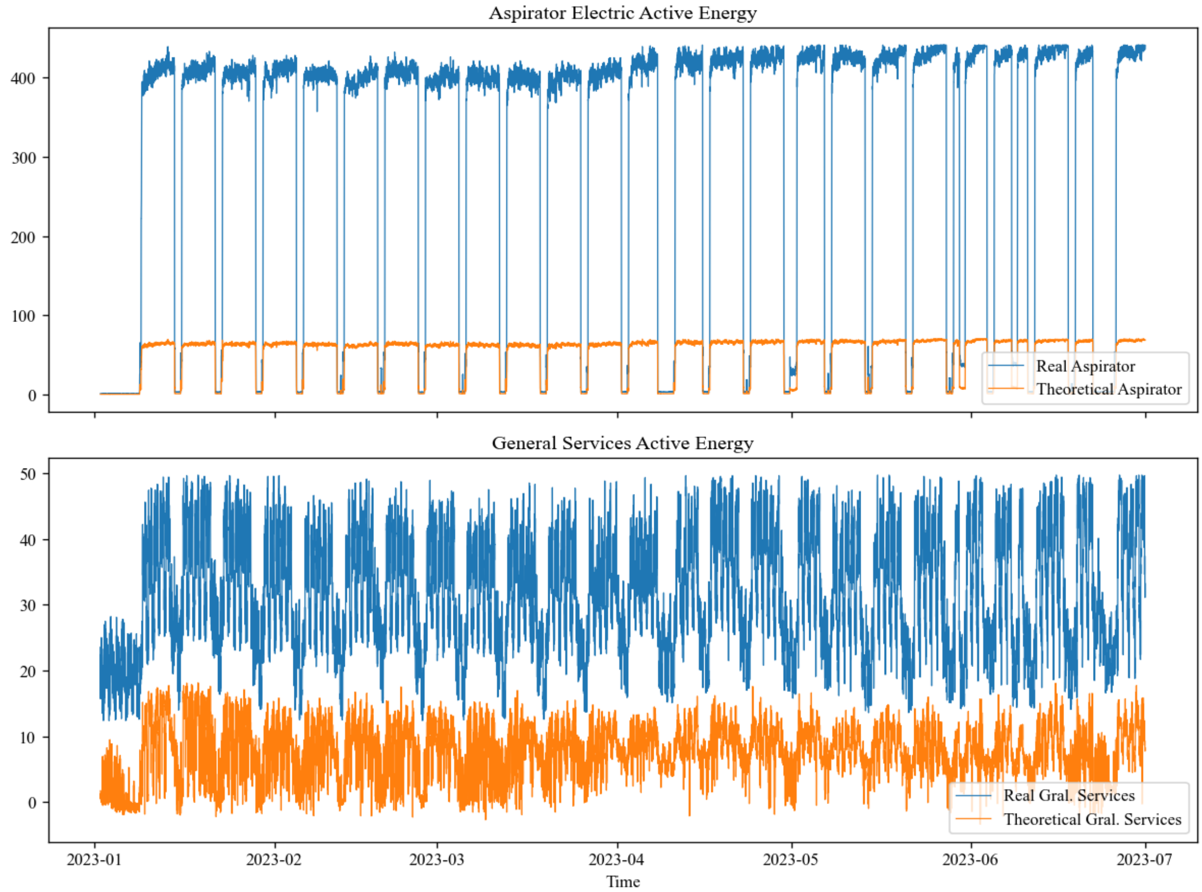



Figure 2: Differences between the Active energy obtained from the smart sensors vs the predicted ones in two sectors for Industry Site 3 for the last 6-month data.

### 5.3.2 Aggregation of Active Energies

In the dataset there were in the range of 10's variables and then a variable usually referred to as `'General Electric Active Energy'`. The team thought that the latter would be the sum of the others, as we aimed to find in Figure 3. But as it happened with the previous case, it turned out not to have any sort of consistency between datasets. So, to be conservative, it was assumed there was no correlation between each individual sector and the General variables.

### 5.3.3 Hourly Aggregation

A subsequent realization was that, since all necessary variables (particularly, energy and power), are scalar quantities with only magnitude, the individual measurements can be summed, aggregating these magnitudes hourly. This approach provides several benefits:

1. By aggregating 15-minute interval data into hourly data, the size of the dataset was reduced by a factor of four. This significant reduction makes the dataset more manageable and easier to handle in computational processes.
2. When relating the variables to other parameters or decision variables, it would be easier to match them since these are frequently presented as hourly data. For example, energy prices or weather data come in hour granularity.

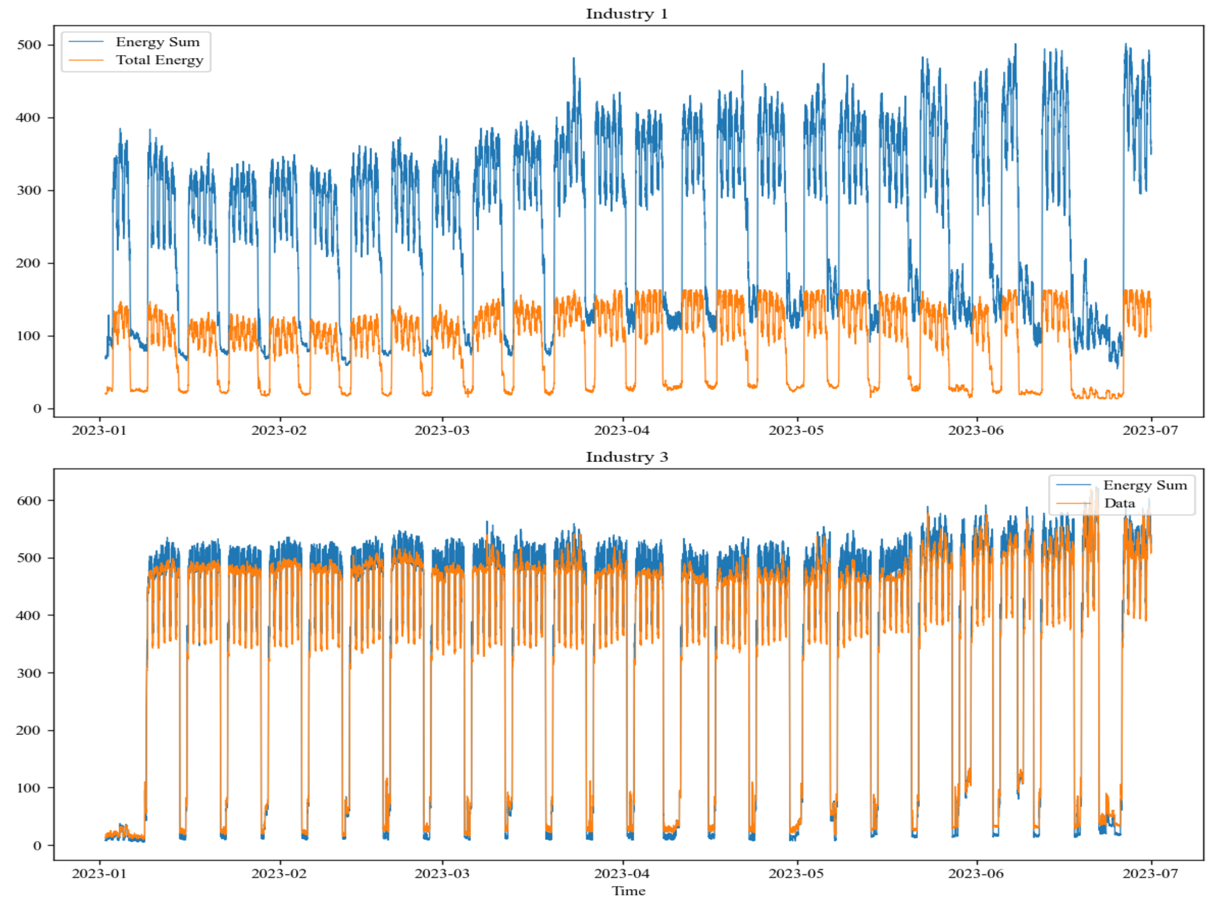



Figure 3: Comparison for Industry Site 1 & 3 between the General Active Energy and the sum of the individual sectors.

## 5.4 Processed Industry Site 2

At this point it is important to mention the reasons why the team decided to proceed with a certain dataset, given the information already acquired in the exploratory data analysis.

As mentioned in subsection 5.2, Industry Site 2 exhibits a significantly higher number of null values compared to the other two datasets, differing by two orders of magnitude. This difference initially led us to focus on an analysis for Industry Sites 1 & 3. However, after studying the different elements that take into account as stated in section 4, the focus changed to conduct the optimization study for a processed Industry Site 2 dataset. Despite the presence of null values, there are other reasons that had more weight towards this decision, which are listed below:

1. The variables presented in Industry 2 seemed more interesting towards the creation of constraint equations. They permitted a more versatile optimization analysis.
2. After studying how Pyomo would process the constrain equations, and based on the experience of experts in this field, it was recommended that we perform a simultaneous analysis of no more than 10 constraint equations. This rule of thumb is guided by the notion that a larger number of equations would lead towards heavy numerical computational resources.

Therefore, Industry Site 2 was selected to continue this study[5]. Moreover, the dataset was modified to satisfy the needs for further work with the optimization algorithms. Such changes are presented below:

### 5.4.1 Updated variables

First, the new processed dataset contained information regarding each unique sector. This are: `Production, Chiller, UTA, Compressor, Offices and Data Centers`. In addition, several features introducing external conditions have been introduced. The modifications resulted in a dataset with the following format:

- Industrial Sites:
  - Active Energy and Power: As in the original dataset measurement, the new one contains the hourly aggregation of these two magnitudes. Since the aggregation is hourly, Power and Active Energy express the same magnitude (kWh).
  - Power factor: Already present in the original dataset.
  - Apparent Power: Apparent (kVA) was added for completion of the physical landscape of the dataset.
- External Inputs:
  - Seasonality: features like whether it is weekend or not, day of the week.
  - Environmental Conditions: external temperature, precipitations, wind speed, atmospheric conditions where introduced to have the possibility of being used. They were introduced through the Visual Crossing API.
  - Pricing: By using OMIE data, the hourly electric energy price of the Spanish market was used as pricing option.
- Synthesized Data:

Due to lack of on-site data collection, data was synthesized to simulated real world energy consumption from different parts of the industrial site. Data was synthesized with a variety of methods depending on the sector being simulated.
Chiller's data was simulated using real-world relationship between heat index, efficiency, a set point rules. Data center was simulated by taking know hardware consumption values and approximating to known consumption values using `SciPy`. Other sectors were simulated by calculating a scale factor over the max consumption of that sector, multiplying this factor to an expected value for the variable, and generate points with normal distribution on the expected value.

  - Data Center:
    * Number of Active Servers
    * Number of Active PoE Network Switches
    * Number of Active non-PoE Network Switches
    * Number of Active HDDs

[5]The conclusion extracted up to this point from the exploratory data analysis for Industry Sites 1 & 3 are perfectly applicable so far and do not contradict the further work on the new dataset.

    * Number of Active SSDs
  - Chiller Group:
    * Temperature Set Point
    * Number of Active Chillers
  - Production:
    * Number of Workers
    * Power Used to Charge Transport Vehicles
    * Production Schedule
    * Maintenance Status
    * Volume Production Waste
  - UTA (Ultrasonic Testing Apparatus):
    * Operational Presence
    * Fabric in Chamber
    * Testing Schedule
    * Workload
    * Standby Power Down
  - Office:
    * Number of Active Light Bulb
    * Number of Active Wall Plugs
    * Number of Active Computers
    * Number of Active Printers
    * Number of Active Coffee Machines

### 5.5 Nulls and Outliers

Overall, nulls were identified and replaced using linear interpolation. This is used to reflect the temporal nature of a data. Using this method every missing value between two non null values are replace with the outcome of a linear equation between the two values. Outliers were handled by identifying values outside of $1.5{\cdot}$IQRs and imputing them with the feature's median.

## 6 Results

With the new processed dataset, and understanding how each feature takes action, it was time to run the optimization algorithm. For that, the modeled constraint equations for each sector within the industry are presented below. Each will be a combination of decision variables, parameters and its bounds. Parameters are introduced to the optimizer for each hour of the day and the optimizer solves to minimize the objective function.

## 6.1 Constrain Equations

**Production:** The model quantifies how different operational factors influence the total energy consumption within a production facility.

$$P = \alpha_0 + \alpha_1 \cdot \mathrm{PTV} + \alpha_2 \cdot \mathrm{PS} + \alpha_3 \cdot \mathrm{MS} + \alpha_4 \cdot \mathrm{VPW} + \alpha_5 \cdot \mathrm{NW} \tag{5}$$

The formula represents a model where $P$ quantifies an output based on several input variables, each weighted by a coefficient represented as $\alpha$. Specifically:

- $\alpha_1$ multiplies PTV (Power Transport Vehicles), indicating the energy attributed to charging operation vehicles.
- $\alpha_2$ is associated with PS (Production Schedule), reflecting how production activity levels impact the output.
- $\alpha_3$ corresponds to MS (Maintenance Status), showing the influence of equipment maintenance on the outcome.
- $\alpha_4$ is tied to VPW (Volume Production Waste), which accounts for waste management efficiency.
- $\alpha_5$ multiplies NW (Number of Workers), representing the effect of workforce size on the result.

**Chiller:** The model quantifies how different operational factors influence the adjusted energy consumption of the chiller system.

$$C = \beta_0 + \beta_1 \cdot \mathrm{HISPD} + \beta_2 \cdot \mathrm{EAP} + \beta_3 \cdot \mathrm{NAC} \tag{6}$$

- $\beta_1$ multiplies HISPD (Heat Index Set Point Difference), adjusting energy consumption based on the difference between the heat index and the set point, which might reflect external thermal load or environmental conditions.
- $\beta_2$ is associated with EAP (Efficiency Adjusted Power), reflecting the adjusted power consumption based on the efficiency of the chiller.
- $\beta_3$ corresponds to NAC (Number of Active Chillers), indicating the effect of the operational capacity of chillers on energy consumption.

**UTA:** This formula allows for the management of the Ultrasonic Testing Apparatus (UTA)'s energy consumption.

$$\mathrm{UTA} = \gamma_0 + \gamma_1 \cdot \mathrm{OP} + \gamma_2 \cdot \mathrm{FC} + \gamma_3 \cdot \mathrm{TS} + \gamma_4 \cdot \mathrm{W} + \gamma_5 \cdot \mathrm{SPD} \tag{7}$$

- $\gamma_1$ multiplies OP (Operational Presence), indicating whether the UTA is actively in operation, directly impacting energy use.
- $\gamma_2$ is associated with FC (Frabic in Chamber), reflecting whether fabric is loaded in the testing chamber, suggesting readiness for use.

- $\gamma_3$ corresponds to TS (Testing Schedule), representing planned testing times which predict active usage.
- $\gamma_4$ is tied to W (Workload), measuring the queue of testing tasks assigned to the UTA, affecting how continuously the device operates.
- $\gamma_5$ multiplies SPD (Standby Power Down), managing energy use when the UTA is not actively testing, thereby controlling idle power consumption.

**Office Space:** The model quantifies how different electrical devices influence the total energy consumption within an office setting.

$$OFF = \delta_0 + \delta_1 \cdot \mathrm{LA} + \delta_2 \cdot \mathrm{AWP} + \delta_3 \cdot \mathrm{AC} + \delta_4 \cdot \mathrm{AP} + \delta_5 \cdot \mathrm{ACM} \tag{8}$$

- $\delta_1$ multiplies LA (light bulbs active), reflecting energy usage by the number of active lightbulbs.
- $\delta_2$ is associated with AWP (Active Wall Plugs), indicating energy consumption by devices plugged into active wall outlets.
- $\delta_3$ corresponds to AC (Active Computers), measuring the energy consumption by computers that are currently in use.
- $\delta_4$ is tied to AP (Active Printers), representing the energy usage by printers during the measurement period.
- $\delta_5$ multiplies ACM (Active Coffee Machines), accounting for the energy consumed by coffee machines in use.

**Data Center:** This model quantifies how different hardware components influence the total energy consumption in a data center.

$$\mathrm{DC} = \epsilon_1 \cdot \mathrm{S} + \epsilon_2 \cdot \mathrm{NS\text{-}P} + \epsilon_3 \cdot \mathrm{NS\text{-}NP} + \epsilon_4 \cdot \mathrm{HD} + \epsilon_5 \cdot \mathrm{SSDs} \tag{9}$$

- $\epsilon_1$ multiplies S (Servers), reflecting energy usage by active servers.
- $\epsilon_2$ is associated with NS-P (Network Switches (PoE)), indicating energy consumption by Power over Ethernet switches.
- $\epsilon_3$ corresponds to NS-NP (Network Switches (non-PoE)), measuring the energy consumption by non-PoE network switches.
- $\epsilon_4$ is tied to HHD (Hard Drives), representing the energy usage by active hard disk drives.
- $\epsilon_5$ multiplies SSDs, accounting for the energy consumed by solid-state drives.

## 6.2 Financial Savings

With the final elements presented that take place in the optimization algorithm, it became time to elaborate on the obtained energy savings. The complete workflow regarding the optimization algorithm can be seen in Appendix B.

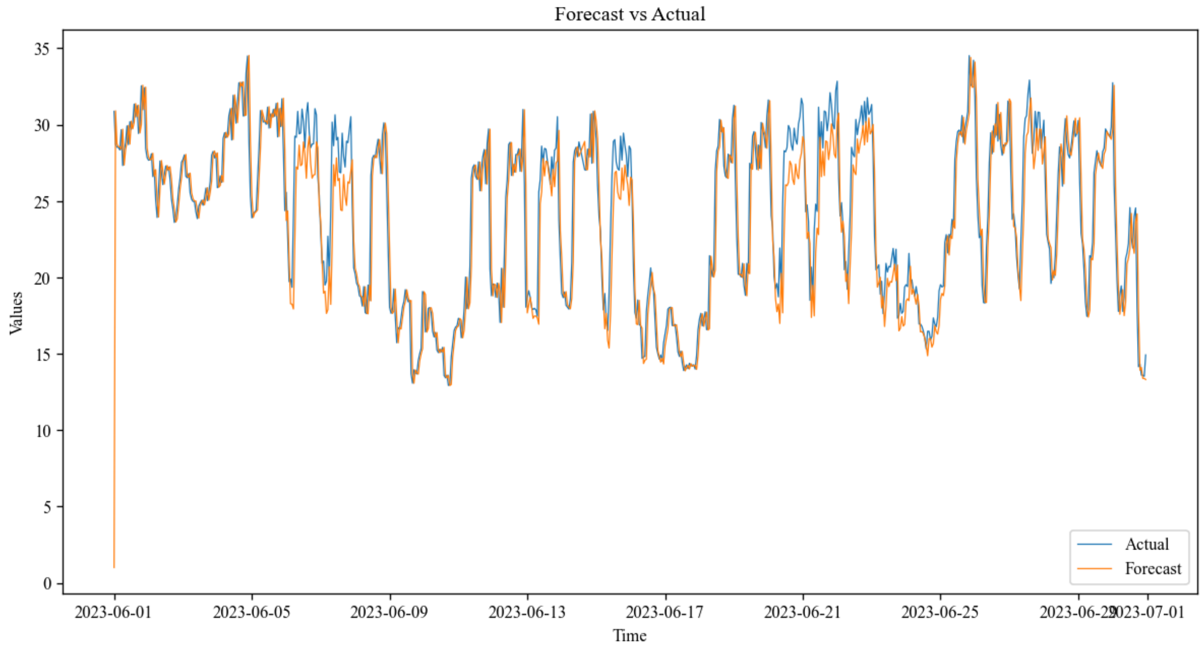



Figure 4: Comparison for the forecasting process using ARIMA(5,1,0) for the Technological Center.

As stated in subsection 4.5, forecasting was essential for both decision variables and parameters. In Figure 4 it is shown the capabilities to forecast based on historical data for the technological centers.

In the month of June 2023, the optimizer outputs a value of 24,660.03 € while the ground truth had a value of 31,035.43 €. Therefore, with a difference value of **6,375.40 €**, the optimizer resulted in savings by 20.5% as presented in Figure 5.

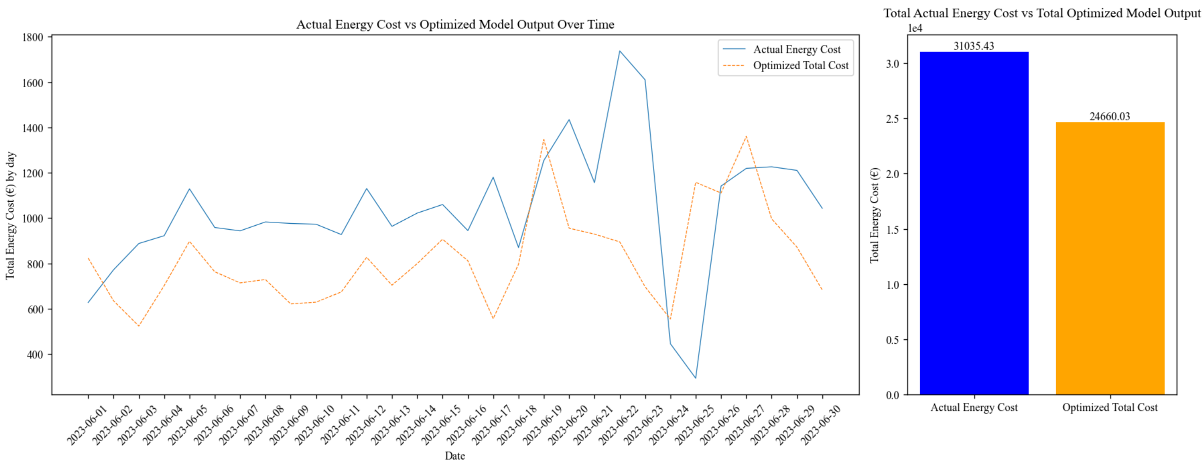



Figure 5: Comparison between the ground truth cost and optimizer results for Industry 2 for the month of June 2023. Optimization algorithm proved to achieve savings.

# 7 Conclusions

In conclusion, this Capstone Project has effectively demonstrated the culmination of skills acquired throughout this Master's program and Venture Lab. However, it's essential to acknowledge certain challenges and outcomes of this work.

First, working with an online dataset posed barriers to fully comprehending the operational intricacies of the industrial site. Physical presence in the industry could have provided deeper insights and potentially improved the accuracy of assumptions and analysis. Secondly, we can not fully guarantee that the variables present represented the totality of sectors within the Industry Site; which could explain the discrepancies experienced during the EDA. Furthermore, the continuous runs of inexplicably empty periods and outlier values could be explained if we did not assume the correct operation of the sensor.

In addition, one big improvement in this work was the cleaning of the present datasets and the later processing to include external information like energy prices, weather conditions and the introduction of synthesized data.

Furthermore, the optimizer has proved effectiveness. As in most optimization cases consulted, the average improvement is around $10 \pm 5\%$. So, Electritect has demonstrated its capabilities by achieving savings on approximately double this factor. In conclusion, the team considers this work to have an above average and exceptional outcome.

## 7.1 Future Development

To conclude, we would like to put forth other lines of work that Electritect would like to explore in the near future.

Other sources of optimization software and techniques, along with more sophisticated equations for the forecasting processes would like to be addressed at first.

Also, such implementation would like to be as standardized as possible, which would allow our clients to reduce implementation time and drive value fast. This would include models to be stored in the could and the exploration of techniques like `Q-Learning`, which is a reinforcement learning algorithm that learns quality of actions telling an agent what action to take under what circumstances.

Moreover, an in-depth orientation period would need to take place to ensure all sources of energy consumption are accounted for the optimization model. In relation to this, additional data would need to be collected from real world examples in order to avoid using synthesized data in production models.

We would like to explore other sources of energy usage like natural gas or water system, which also have associated costs that can be minimised.

An avenue that our team considered was developing the hardware and software ourselves to monitor and manage electricity usage remotely. We explored a hardware configuration of ESP32 microcontroller, SCT-013-050 current sensor and ZMPT101B voltage transformer to measure real-time energy consumption. The ESP32 would connect to the internet via the built in WiFi and transmit the data to the Blynk 2.0 platform.

Although we were successful in building a prototype IoT hardware and software solution, we still believe it would be beneficial to explore partnerships with developed IoT suppliers. Our solution was an exercise in electrical and data engineering but would be near impossible to scale with the resources and knowledge at our disposal.

To finalize, building on the project's achievements and the future improvements, the team anticipates further increasing cost savings by an additional 10% with these future developments.

# References


[1] Hannah Ritchie, Pablo Rosado, and Max Roser. "Access to Energy". In: *Our World in Data* (2019). URL: https://ourworldindata.org/energy-access.

[2] Antonio Gomez-Exposito. *Análisis y operación de sistemas de energía eléctrica*. Jan. 2002. ISBN: 84-481-3592-X.

[3] World Economic Forum. "Global Energy Transition and its Implications on Major Industrial Sectors". In: (June 2024).

[4] United Nations Climate Change. "The Paris Agreement and its Implications for Global Industries". In: (2021). URL: https://unfccc.int/process-and-meetings/the-paris-agreement..

[5] International Energy Agency Reports. "Energy Policies and Compliance for Industrial Sectors". In: *International Energy Agency Reports* (2023). URL: https://www.iea.org/energy-system/industry.

[6] McKinsey & Company. "Managing Industrial Energy Expenses: Trends and Strategies". In: *McKinsey & Company Insights* (2023).

[7] EU Commission. "The European Green Deal and Its Impact on Industry". In: *European Commission Reports* (2022).

[8] Abdul Salam Shah et al. "A Review on Energy Consumption Optimization Techniques in IoT Based Smart Building Environments". In: *Information* 10.3 (2019). ISSN: 2078-2489. DOI: 10.3390/info10030108. URL: https://www.mdpi.com/2078-2489/10/3/108.

[9] Henning Kagermann et al. *Recommendations for implementing the strategic initiative INDUSTRIE 4.0: Securing the future of German manufacturing industry; final report of the Industrie 4.0 Working Group*. Forschungsunion, 2013.

[10] Michael E Porter, James E Heppelmann, et al. "How smart, connected products are transforming competition". In: *Harvard business review* 92.11 (2014), pp. 64–88.

[11] Jaime Lloret et al. "An integrated IoT architecture for smart metering". In: *IEEE Communications Magazine* 54.12 (2016), pp. 50–57.

[12] Thomas Boyle et al. "Intelligent metering for urban water: A review". In: *Water* 5.3 (2013), pp. 1052–1081.

# A Other datasets

Here we attach a few datasets that, while not satisfying the total conditions for the project, seemed interesting to take into account

- 7-Story Office Building in Bangkok, Thailand
- Buildings Energy Consumption
- Energy and Water Data Disclosure for NYC
- 3-year Building Operational Performance

# B Process Workflow

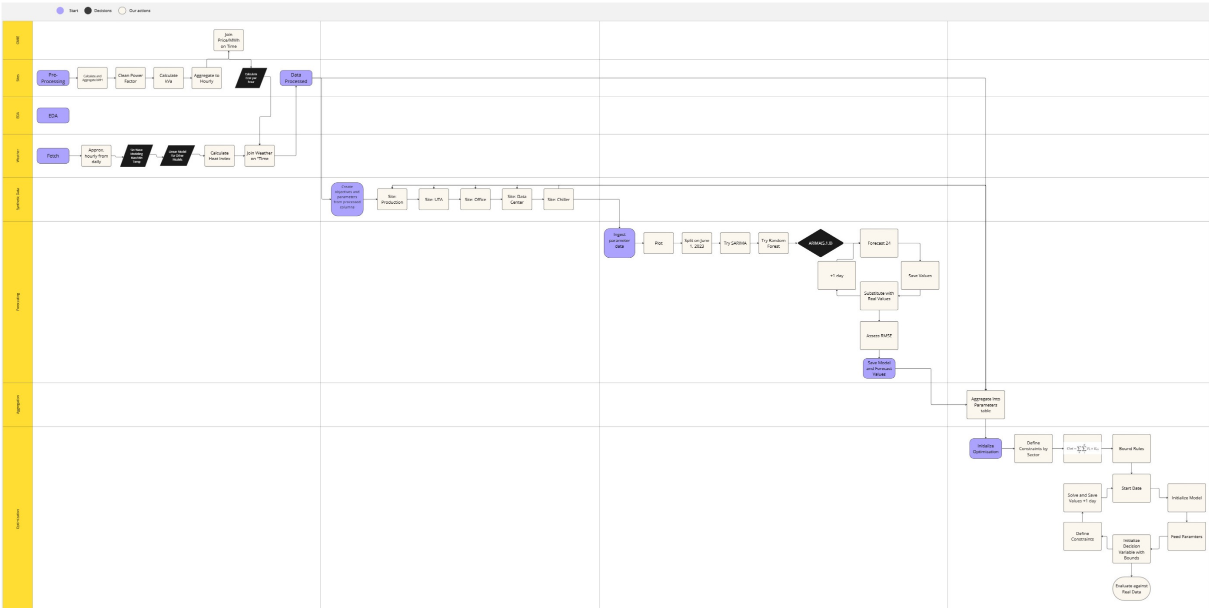


Figure 6: Complete process workflow related the optimization algorithm. From data ingestion to output results.

# C Industrial Site Variables

## C.1 Industrial Site 1

['Time', '01 Normal General Electric Active Energy', '02 Chiller Electric Active Energy', '03 Dyeing Electric Active Energy', '04 Ironing Electric Active Energy', '05 Purifier Electric Active Energy', '07 Technological Centers Electric Active Energy', '08 Offices Changing Rooms Electric Active Energy', '09 Compressed Air Electric Active Energy', '10 General Services Electric Active Energy', '11 UPS Electric Active Energy', '12 UTA Electric Active Energy', '14 Laboratory Electric Active Energy', '15 Warehouses Electric Active Energy', '17 Print Electric Active Energy', 'General Electric Active Energy', 'General Natural Gas Quantity', 'General Vapour Quantity', 'General Vapour Fumes Temperature', '01 Normal General Electric Power Factor', '02 Chiller Electric Power Factor', '03 Dyeing Electric Power Factor', '04 Ironing Electric Power Factor', '05

Purifier Electric Power Factor', '07 Technological Centers Electric Power Factor', '08 Offices Changing Rooms Electric Power Factor']

## C.2 Industrial Site 2

['Time', '01 General Electric Active Energy', '02 Production Electric Active Energy', '03 Chiller Group Electric Active Energy', '04 UTA Electric Active Energy', '05 Compressors Electric Active Energy', '06 Offices Electric Active Energy', '07 Data Center Electric Active Energy', '08 Technological Centers Electric Active Energy', 'General Electric Active Energy', '01 General Electric Power Factor', '02 Production Electric Power Factor', '03 Chiller Group Electric Power Factor', '04 UTA Electric Power Factor', '05 Compressors Electric Power Factor', '06 Offices Electric Power Factor', '07 Data Center Electric Power Factor', '08 Technological Centers Electric Power Factor', 'General Electric Power Factor', '01 General Electric Active Power', '02 Production Electric Active Power', '03 Chiller Group Electric Active Power', '04 UTA Electric Active Power', '05 Compressors Electric Active Power', '06 Offices Electric Active Power', '07 Data Center Electric Active Power', '08 Technological Centers Electric Active Power', 'General Electric Active Power', 'Unnamed: 28']

## C.3 Industrial Site 3

['Time', '01 General Transformer 1234 Electric Active Energy', '02 Chiller Group Electric Active Energy', '03 Aspirator Electric Active Energy', '04 Compressed Air Electric Active Energy', '05 Weaving Electric Active Energy', '06 Ironing Electric Active Energy', '07 UPS Electric Active Energy', '08 General Services Electric Active Energy', '09 UTA Electric Active Energy', '10 Warehouses Electric Active Energy', '11 Winding Electric Active Energy', '12 Others Electric Active Energy', 'General Technical Efficiency', '01 General Transformer 1234 Electric Power Factor', '02 Chiller Group Electric Power Factor', '03 Aspirator Electric Power Factor', '04 Compressed Air Electric Power Factor', '05 Weaving Electric Power Factor', '06 Ironing Electric Power Factor', '07 UPS Electric Power Factor', '08 General Services Electric Power Factor', '09 UTA Electric Power Factor', '10 Warehouses Electric Power Factor', '11 Winding Electric Power Factor', '12 Others Electric Power Factor', 'General Technical Pressure', 'General Technical Temperature', '01 General Transformer 1234 Electric Active Power', '02 Chiller Group Electric Active Power', '03 Aspirator Electric Active Power', '04 Compressed Air Electric Active Power', '05 Weaving Electric Active Power', '06 Ironing Electric Active Power', '07 UPS Electric Active Power', '08 General Services Electric Active Power', '09 UTA Electric Active Power', '10 Warehouses Electric Active Power', '11 Winding Electric Active Power', '12 Others Electric Active Power', 'General Technical Active Power', 'General Technical Flow Rate']