Optimization of CO_2 Emission Reduction of an Industrial Plant Through Solar Self-consumption and Electrical Storage Improvement

Table of contents

A. Introduction to the Challenge

The energy transition in Spain is progressing steadily, but it currently faces a critical challenge: **how to efficiently integrate the enormous amount of renewable energy being deployed?**

According to the **National Integrated Energy and Climate Plan (PNIEC) 2023**, by 2030 Spain will have **160 GW of installed renewable capacity**, of which **138 GW will be wind and solar**. However, this expansion faces an inescapable reality: **renewables are intermittent, and the electrical system cannot always absorb all the energy they generate**.
In 2022, 3.3 TWh of wind energy were wasted because it could not be used at the right time. The lack of synchronization between **production** and **consumption** is causing energy curtailments and, paradoxically, the need to continue relying on fossil sources to ensure supply during periods of low renewable generation.

To solve this issue, **Spain needs flexibility mechanisms** that allow for intelligent demand management and better utilization of available renewable energy. This is where **self-consumption systems and their hybridization with batteries, especially in the industrial sector, can make a difference**.

Many industries have installed photovoltaic self-consumption systems with no feed-in to the grid, but they do not always take full advantage of the energy generated. In this datathon, **we will analyze an industrial installation of 175 kWp** that has detected that it could have used more electricity from its solar panels over a year of analysis (July 2023 – August 2024).

The challenge in this datathon is twofold: first, to calculate the **maximum generation potential** that the plant could have achieved based on weather data; and second, to **optimize the use of a 'theoretical' battery to jointly maximize the reduction of CO_2 in its energy consumption through self-consumption and battery usage** for the period of September 2024.

B. Datathon objectives

Participants must develop models that allow for:

1. **Calculation of solar generation potential**:

   • Estimate the **maximum possible generation during the forecast period** based on weather data, installation characteristics, and photovoltaic generation produced.
   • Compare the energy actually used with the energy that could have been utilized. This happens when, for example, in a given hour, photovoltaic generation potential would have been 100kWh, but at that time there isn´t enough industrial demand to consume all the solar energy. So if industrial demand during that hour is only 40kWh, but there is a solar generation of 100kWh, there would be 60kWh of solar generation that could have been utilized but instead was lost/wasted.

2. **Optimization of battery storage**:

- **Simulate the use of a "typical" 100 kWh battery (capacity) and a maximum charge/discharge power of 100 kW, allowing only one charge/discharge cycle per day.**

  Given these battery characteristics, the battery could, for example, charge at 100kW for 1 hour (100kWh) or similarly, charge 40kWh in the first hour, another 40kWh in the second hour and 20kWh in the third hour. The amount of energy that can be charged at a given hour will depend on the excess solar generation, and the amount of energy that can be discharged in a given hour will depend on how much the factory needs to consume at a given hour.

  This battery is disconnected from the grid, so the energy stored in the battery can only be used by the factory.

- **Maximize self-consumption and storage while reducing grid dependence** in periods when the **energy mix has a higher proportion of high CO_2-emitting sources**.

3. **Environmental impact analysis**:

   - Compare the carbon footprint (CO_2eq emissions) of the current situation without optimization VS **the simulated scenario with the use of the battery**.

The forecast period is **September 2024**, and you will be provided with **photovoltaic generation (kWh)** and **energy consumption from the grid (kWh)** (hourly data) covering the period from **July 24, 2023, to August 31, 2024**. You will be also provided with the actual photovoltaic energy consumption/use of the factory during the period of September 2024.

Additionally, **meteorological data** (hourly basis) will be available from **July 24, 2023, to October 31, 2024**.

Furthermore, the **carbon intensity of the Spanish electricity mix** (gCO_2eq/kWh, hourly data) will be provided for the **entire years 2023 and 2024**.

1. Objective 1: Calculation of Solar Generation Potential

The factory has a solar installation of 175 kWp of self-consumption capacity with no feed-in to the grid (1:1 ratio with the nominal inverter power). However, during the forecast period, it was observed that the factory did not use all the electricity it could have generated (weekends consumed very little, sometimes due to poor maintenance, unresolved technical problems, etc.).

The goal of this first part of the challenge is to **quantify how much additional energy could have been used if it had been optimally managed** (it is important to note that since it cannot feed excess generated energy back to the grid, it is lost).

The objective here is to develop a model that estimates the **maximum possible solar generation at each moment on an hourly basis**, considering weather conditions and historical generation data for the analysis period—as well as the characteristics of the photovoltaic installation.

To reach the optimal result, the ideal use of **meteorological data** must be analyzed (i.e., averaging the distance to the c**oordinates of the 4 closest points** with data relative to the factory's potential location, or other types of analysis, which meteorological variables should be integrated, etc.).

- **First step of this objective 1:** develop a model that estimates the **maximum possible solar generation at each moment on a hourly basis** for one month (September 2024). The output file will contain two fields: Date and time (hour) and energy generated. **MAE** will be the performance metric with at least four decimal metrics. You can use **"**the calculator**"** to check your MAE.

Field Name	Description
Datetime	Date and time of the energy generation
KWH_ENERGIA	Total photovoltaic energy generated (in kWh).

- **Second step of this objective 1:** Then, compare this potential generation with the actual energy generated (the amount of solar generation that was actually consumed by the factory) for September 2024 in kWh, to **calculate the underutilization of the installation.** You can use **"**the calculator**"** to calculate it. **Note:** The output will consist solely of the total calculation for the period, presented as a single figure with four decimal places (in kWh).

2. Objective 2: Optimization of Battery Usage in the Installation.

The goal is to **maximize self-consumption** by utilizing a theoretical battery to store **excess solar energy** and subsequently use it when needed. This approach reduces **dependence on the electrical grid**, prioritizes the use of **locally generated renewable energy**, and contributes to the **reduction of CO_2 emissions**.

Using your prediction of the total photovoltaic energy generated (in kWh) for September, and with knowledge of the actual photovoltaic energy consumption/use of the factory during the period of September 2024 (File 5: 'Consumido Fotovoltaica'), you can calculate the daily and hourly surplus of photovoltaic energy. This surplus can then be stored and used later to contribute to the reduction of CO_2 emissions.

- **First step of the objective 2:** To measure the level of self-consumption achieved, the **Ra (Self-Consumption Ratio)** metric is defined. This metric quantifies the percentage of **potentially generable energy** that has been **effectively utilized**, either **directly** or **through battery storage**.

$$R_a = \frac{\text{Photovoltaic energy used directly} + \text{Energy stored and subsequently used}}{\text{Potentially generable energy}} \times 100$$

To achieve this, the battery's behaviour throughout the forecast period must be **simulated**, optimizing **charge and discharge cycles** based on **solar generation, energy demand, and the existing energy mix**.

Key considerations:

1. Use a theoretical battery with predefined characteristics.

- Capacity: 100 kWh
- Maximum charge/discharge power: 100 kW
- Allowing only one charge/discharge cycle per day

2. **One charge/discharge cycle per day** means that the battery can charge fully (either in the same hour or over several hours until its fully charged at 100kWh, depending on the excess energy available to charge) and then discharge fully (again, over one hour or over several hours until it fully discharges the 100kWh, depending on the amount of energy being consumed from the grid at a given hour) only once in a given day.

3. **The battery is disconnected from the grid** and can only be charged with excess solar energy.
4. If the **battery fills too quickly**, solar energy may be lost due to a lack of available storage capacity. Once the battery reaches full capacity, it can no longer store additional energy, leading to wasted surplus solar generation. Therefore, the charge and discharge timing directly affects self-consumption efficiency and the amount of energy that can be utilized. To optimize battery usage, teams must **model the charging profile** based on the available solar surplus. This involves **using generation and consumption data to decide when and how much energy will be available for charging**. **Additionally, teams should discharge the stored energy at the hour or hours of each day with the highest carbon intensity being consumed from the grid.**

5. **A higher Ra indicates better utilization of the generated solar energy. For this part of the challenge you will be evaluated according to this Ra metric, where the higher the values result in a better score.**

The output will consist of:

- **Calculation of Ra**, which will be presented as a single figure (%) with four decimal places, representing the total calculation for the period. **No performance metric is associated with this calculation.**

- **Assessment of the methodology and calculations, which will be evaluated based on the quality of the approach used.** A detailed explanation of the calculation process must be included in the executive summary. You are required to provide a comprehensive and well-documented explanation of the methodology applied, specifying the assumptions, criteria, and computational steps used to obtain the results. Clear and structured documentation will be a key factor in the evaluation.

3. Objective 3: Environmental Impact Analysis

Finally, the CO_2 reduction derived from increased solar energy usage and storage, as opposed to grid energy, **must be calculated in periods with a high-emission energy mix**. The optimal use of solar energy in an industrial installation, combined with storage, reduces grid dependence, which in turn decreases CO_2 emissions associated with electricity consumption.

This last part of the challenge is the **straightforward** and aims to **quantify the CO_2 emissions that could have been avoided** if the installation had **better utilized its photovoltaic generation and had access to battery storage**. Since **economic performance optimization is not the goal** of this challenge, this value should represent the **maximum achievable CO_2 reduction** for a facility of these characteristics during the analysis period (September 2024).

To quantify this, we define the **Total CO_2 Reduction** $CO2_{ev}$, calculated as the difference between:

1. **CO_2 emissions in the baseline scenario** (without battery optimization).
2. **CO_2 emissions in the optimized scenario** (with battery storage optimization).

$$CO2_{ev} = CO2_{emitted} - CO2_{emitted}^{optimized}$$

Key considerations:

1. $CO2_{emitted}$ → CO_2 emissions based on **actual grid consumption**. This represents the total **CO_2 emissions under normal operation**, without battery optimization. The summation **adds up the CO_2 emissions for each hour over the entire period.**

$$CO2_{emitted} = \sum (\text{Energy consumed in each hour} \times \text{CO}_2 \text{ intensity in each hour})$$

2. $CO2_{\substack{optimized \\ emitted}}$ → CO_2 emissions after applying an optimized battery storage strategy. The summation adds up the new, **optimized CO_2 emissions for each hour**.

$$CO2_{emitted}^{optimized} = \sum (\text{Optimized energy consumption in each hour} \times \text{CO}_2 \text{ intensity in each hour})$$

3. For this part of the challenge, your evaluation will be based on the $CO2_{ev}$ metric, where higher values indicate a more effective emissions reduction strategy. **The greater the value, the higher the score.**

The output will consist of:

- **Calculation of $CO2_{ev}$**, is expressed in grCO_2 equivalent avoided. It must be presented as a single figure with four decimal places, representing the total calculation for the period. **No performance metric is associated with this calculation.**
- **Assessment of the methodology and calculations, which will be evaluated based on the quality of the approach used.** A detailed explanation of the calculation process must be included in the executive summary. You are required to provide a comprehensive and well-documented explanation of the methodology applied, specifying the assumptions, criteria, and computational steps used to obtain the results. Clear and structured documentation will be a key factor in the evaluation.

For Objectives 2 and 3, you must generate an output file that contains the following information:

Field Name	Label	Description
Datetime	**Datetime**	Timestamp (Day & Hour) indicating when the energy was stored or discharged.
CO2_Avoided	**Total CO_2 Avoided ($grCO_2$ eq)**	Total CO_2 (**$grCO_2$ eq**) emissions avoided by using stored energy instead of drawing from the grid. CO_2 emitted - CO_2 emitted optimized

C. Datasets

The **forecast period** is **September 2024**, and you will be provided with the following information:

1. **Photovoltaic generation** (kWh, hourly data) – Available from **July 24, 2023, to August 31, 2024**.
2. **Energy consumed from the grid** (kWh, hourly data) – Available from **July 24, 2023, to August 31, 2024**.

These three datasets include information for the **forecast period**, as they are required for conducting the necessary analysis and predictions:

3. **Meteorological data** from the four nearest points to the plant (15 different variables on an hourly basis) – Available from **July 24, 2023, to October 31, 2024**.
4. **Carbon intensity of the Spanish electricity mix** (gCO_2eq/kWh, hourly data) – Available for the **entire years of 2023 and 2024**.
5. **Photovoltaic energy consumption/use:** Actual photovoltaic energy consumption/use of the factory during the period of September 2024

This file contains useful information/examples for you to understand the business case:

6. **Example Batery optimization:** This Excel file **contains a conceptual example** of how to carry out the optimization processes for both charging and discharging the theoretical battery, following the criteria defined in this challenge.

File 1 Description: "Generacion_fotovoltaica"

This file contains data on photovoltaic energy generation from the factory located in Pinto. It includes timestamps indicating when the measurements were taken and the corresponding amount of energy generated in kilowatt-hours (kWh). The dataset can be used to analyze energy production trends, compare generation with weather conditions, and optimize self-consumption strategies in industrial installations. Hourly data. This dataset includes entries where some values are equal to zero. Covers the period from July 24, 2023, to August 31, 2024.

VARIABLE	DESCRIPTION
FECHA	Date and time of photovoltaic generation in Europe/Madrid time zone.
TOTAL_KWH_ENERGIA	Total photovoltaic energy generated (in kWh) in the period of time (Fecha).

Illustrative example:

```
                     TOTAL_KWH_ENERGIA
FECHA
2023-07-24 11:00:00             115.66
2023-07-24 12:00:00             129.98
2023-07-24 13:00:00             134.35
2023-07-24 14:00:00             133.90
```

File 2 Description: "Consumo"

This dataset contains the electrical grid consumption of the Pinto factory with hourly granularity. It covers the period from July 24, 2023, to August 31, 2024.

The dataset provides timestamps and the corresponding energy consumption values (in kWh), allowing for an analysis of **energy demand, optimization of self-consumption**, and evaluation of potential energy storage strategies. Hourly data.

VARIABLE	DESCRIPTION
FECHA	Date and time of the energy consumption measurement in Europe/Madrid time zone.
TOTAL_KWH_ENERGIA	Total energy consumed from the power grid (in kWh) by the factory at each recorded timestamp.

Illustrative example:

```
                     TOTAL_KWH_ENERGIA
FECHA
2023-07-24 11:00:00         506.012634
2023-07-24 12:00:00         456.551453
2023-07-24 13:00:00         395.318604
2023-07-24 14:00:00         330.052979
```

File 3 Description: "Meteorologia"

This file contains weather forecasts obtained from the Global Forecast System (GFS): https://www.ncei.noaa.gov/products/weather-climate-models/global-forecast

It covers the period from July 24, 2023, to October 31, 2024. The file includes a table listing various meteorological variables along with their descriptions. These variables are used in weather forecasting and energy modeling, particularly for analyzing solar energy potential and optimizing battery storage in industrial applications. The dataset includes key meteorological factors such as temperature, humidity, wind speed, solar radiation, cloud cover, and atmospheric pressure, measured at different altitudes and time intervals. Data is provided on an hourly basis.

Variable	Description
FORECAST_TIMESTAMP	Forecast time date in UTC time zone.
LATITUDE	Latitude of the forecast location in geographic coordinates.
LONGITUDE	Longitude of the forecast location in geographic coordinates.
10uheightAboveGround_10	U component of the wind (east-west direction) at 10 meters above the ground in m/s.
10vheightAboveGround_10	V component of the wind (north-south direction) at 10 meters above the ground in m/s.
2rheightAboveGround_2	Relative humidity at 2 meters above the ground in percentage (%).
2shheightAboveGround_2	Specific humidity at 2 meters above the ground in kg/kg.
2theightAboveGround_2	Air temperature at 2 meters above the ground in degrees Kelvin (K).
SUNSDsurface_0	Sunshine duration on the surface in seconds.
aptmpheightAboveGround_2	Apparent temperature at 2 meters above the ground in degrees Kelvin (K).
dlwrfsurface_0	Downward longwave radiation on the surface in W/m^2.
dswrfsurface_0	Downward shortwave radiation on the surface in W/m^2.
gustsurface_0	Maximum wind gusts on the surface in m/s.
msletmeanSea_0	Sea level pressure in Pascals (Pa).
presheightAboveGround_80	Atmospheric pressure at 80 meters above the ground in Pascals (Pa).
pwatatmosphereSingleLayer_0	Total precipitable water in the atmosphere in kg/m^2.
qheightAboveGround_80	Specific humidity at 80 meters above the ground in kg/kg.
spsurface_0	Surface pressure in Pascals (Pa).
tccatmosphere_0	Total cloud cover in the atmosphere in percentage (%).
theightAboveGround_80	Air temperature at 80 meters above the ground in degrees Kelvin (K).
tmaxheightAboveGround_2	Maximum air temperature at 2 meters above the ground in degrees Kelvin (K).
tminheightAboveGround_2	Minimum air temperature at 2 meters above the ground in degrees Kelvin (K).
tozneatmosphereSingleLayer_0	Total amount of ozone in the atmosphere in Dobson Units (DU).
tpsurface_0	Total accumulated precipitation on the surface in mm.
tsurface_0	Surface temperature in degrees Kelvin (K).
uheightAboveGround_80	U component of the wind (east-west direction) at 80 meters above the ground in m/s.
vheightAboveGround_80	V component of the wind (north-south direction) at 80 meters above the ground in m/s.

Illustrative example (only some variables):

```
                           LATITUDE  LONGITUDE  10uheightAboveGround_10  \
FORECAST_TIMESTAMP
2023-07-24 11:00:00+00:00      40.0       -4.0                 6.989717
2023-07-24 12:00:00+00:00      40.0       -4.0                 7.526357
2023-07-24 13:00:00+00:00      40.0       -4.0                 8.506143
2023-07-24 14:00:00+00:00      40.0       -4.0                 9.392205
```

File 4 Description: Carbon intensity of the Spanish electricity mix, measured in grams of CO_2 equivalent per kilowatt-hour (gCO_2eq/kWh), for the years 2023 and 2024. **ES_2023_hourly**; **ES_2024_hourly**

This dataset provides **hourly records of carbon intensity in Spain's electricity mix**. It includes information on:

- **Direct and lifecycle carbon intensity** of electricity generation,
- The **proportion of low-carbon and renewable energy sources** in the electricity mix,
- The **source and reliability** of the data.

This dataset is crucial for **analyzing the environmental impact of electricity consumption**, optimizing energy use in periods of lower emissions, and supporting decisions in renewable energy strategies.

Variable	**Description**
Datetime (UTC)	Timestamp indicating the date and time of the carbon intensity measurement (in Coordinated Universal Time - UTC).
Country	The country for which the data is reported (Spain in this case).
Zone Name	The specific geographic zone within the country (Spain in this case, no subdivisions).
Zone Id	The zone identifier code (e.g., "ES" for Spain).
Carbon Intensity gCO_2eq/kWh (direct)	The direct carbon intensity of electricity generation, measured in grams of CO_2 equivalent per kilowatt-hour (gCO_2eq/kWh). It represents emissions from fossil fuel combustion without considering lifecycle emissions.
Carbon Intensity gCO_2eq/kWh (LCA)	The lifecycle carbon intensity, which includes emissions from fuel extraction, processing, and power plant operations, providing a more comprehensive estimate of the environmental impact.
Low Carbon Percentage	The percentage of electricity generated from low-carbon sources (e.g., nuclear, hydro, solar, wind) relative to the total electricity mix.

Renewable Percentage	The percentage of electricity generated specifically from renewable energy sources (e.g., solar, wind, hydro, biomass).
Data Source	The source of the data (e.g., "entsoe.eu", the European network of transmission system operators).
Data Estimated	A boolean field indicating whether the data is estimated (true) or actual (false).

Illustrative example:

Datetime (UTC)	Country	Zone Name	Zone Id	Carbon Intensity gCOâ,,eq/kWh (direct)	Carbon Intensity gCOâ,,eq/kWh (LCA)	Low Carbon Percentage	Renewable Percentage	Data Source	Data Estimated
01/01/2023 0:00	Spain	Spain	ES	79.49	110.37	84.15	50.75	entsoe.eu	false
01/01/2023 1:00	Spain	Spain	ES	82.59	113.77	83.73	49.24	entsoe.eu	false
01/01/2023 2:00	Spain	Spain	ES	83.2	114.28	83.75	47.23	entsoe.eu	false
01/01/2023 3:00	Spain	Spain	ES	83.55	114.91	83.75	45.71	entsoe.eu	false
01/01/2023 4:00	Spain	Spain	ES	84.06	115.32	83.73	44.1	entsoe.eu	false

File 5 Description: "Consumido_Fotovoltaica"

Actual photovoltaic energy consumption/use of the factory during the period of September 2024.

VARIABLE	DESCRIPTION
FECHA	Date and time of photovoltaic generation in Europe/Madrid time zone.
TOTAL_KWH_ENERGIA	Total photovoltaic energy used (in kWh) in the period of time.

Illustrative example:

A	B
FECHA	TOTAL_KWH_ENERGIA
2024-09-01 09:00:00+02	6.05
2024-09-01 10:00:00+02	20.56
2024-09-01 11:00:00+02	26.05
2024-09-01 12:00:00+02	13.14

File 6: "Example Battery Optimization".

This Excel file **contains a conceptual example** of how to carry out the optimization processes for both charging and discharging the theoretical battery, following the criteria defined in this challenge. **This Excel will assist you in understanding the entire CO2 emission optimization process in a practical manner, using a real-life example.**

In the Excel document you will see a combination of data from different files as well as newly calculated variables:

- **Datetime (UTC)**
- **TOTAL_KWH_ENERGIA (GENERATED)** (your prediction of potential energy generated)

- **TOTAL_KWH_ENERGIA (CONSUMED)** from file consumo_fotovoltaica. This represents ACTUAL SOLAR ENERGY CONSUMED BY THE FACTORY
- **EXCESS_ENERGY_KWH** This is calculated by subtracting actual consumed energy from your prediction of potential energy generated. *It is important to keep in mind that the total sum of the excess energy (solar potential energy generate – actual energy consumed) for the whole month should give you the same result as the value obtained with the calculator for Underutilization of the installation.
- **ENERGY_CHARGED_KWH** This is the amount of excess energy that the battery has used to charge fully
- **ENERGY_LOST** This is the amount of energy that was generated but wasn´t used by the factory and also couldn´t be utilized because the battery was already fully charged, therefore it was wasted.

With the above variables, you will be able to calculated Ra, the Self-Consumption Ratio

$$R_a = \frac{\text{Photovoltaic energy used directly} + \text{Energy stored and subsequently used}}{\text{Potentially generable energy}} \times 100$$

You will calculate: Total Solar Generated, Total Solar Used; Total Energy Not Utilized; Total Energy Recovered with Battery.

Then, to calculate Ra, you would calculate the following: (TOTAL SOLAR USED + TOTAL ENERGY RECOVERED WITH BATTERY)/TOTAL SOLAR GENERATED

- **CONSUMO** from file consume. This is the energy that the factory consumes from the grid.
- **ENERGY_DISCHARGED_KWH** This is the energy that the battery discharged, to satisfy demand from the factory at a given hour
- **REDUCTION_ENERGY_CONSUMED_FROM_GRID** This is the reduction in energy consumed from the grid (CONSUMO – ENERGY_DISCHARGED_KWH). This happens when the factory substitutes clean energy stored in the battery, instead of using battery from the grid at the times when carbon intensity is at its highest.
- **Carbon Intensity gCO2,eq/kWh (direct)** This is the amount, in grams, of CO2 emissions from the electricity coming from the grid. This is at its highest when the electricity coming from the grid is generated using fossil fuels instead of renewable energy
- **CO2_EMITTED**This is the amount of emissions generated, when the factory consumes a given amount of electricity from the grid. If, for example, carbon intensity at a given hour is 100gCO2 per Kwh, and the factory consumes 10kWh from the grid at that time, there would be 100*10 = 100gCO2 emissions
- **CO2_EMITTED_OPTIMIZED** This is the same calculation as above, but you would multiply carbon intensity times REDUCTION_ENERGY_CONSUMED_FROM_GRID. Since the factory is consuming its energy from the battery instead of from the grid, there are less CO2 emissions.

- **TOTAL EMISSIONS** This are the emissions before using the battery. This is the sum of CO2_EMITTED
- **TOTAL CO2 REDUCTION** This is calculated as: CO2_EMITTED - CO2_EMITTED_OPTIMIZED.

D. THE SUBMISIONS: Content to be delivered.

1. Summary of deliveries.

In this section, the contents of the technical deliverables that students must complete to participate in this challenge are described.

Any violation of the delivery criteria will result in the disqualification of the team.

1. **Code must be submitted via the "Final submission Form".**
   The code developed to address the problem (ideally with comments, or in Jupyter Notebooks or Google Colab, alternating between code and text explanations). **The outputs of the execution must be visible.**

2. **Output files containing the predictions/calculations must be submitted via FORMS using the "Check Your Results" button:**

   a. **File 1 (First step of Objective 1): Maximum possible solar generation** at each moment on an hourly basis for September 2024**.** Note that if you make more than one submission, only the last one will be considered for evaluation. Containing the fields: **Datetime and KWH_ENERGIA**

Field Name	Description
Datetime	Date and time of the energy generation
KWH_ENERGIA	Total photovoltaic energy generated (in kWh).

   b. **File 2 (For Objectives 2 and 3):** you must generate an output file that contains the CO2 emissions avoided on an hourly basis for September 2024. Containing the fields: **Datetime and CO2_Avoided.**

Field Name	Label	Description
Datetime	**Datetime**	Timestamp (Day & Hour) indicating when the energy was stored or discharged.
CO2_Avoided	**Total CO_2 Avoided ($grCO_2$ eq)**	Total CO_2 (**$grCO_2$ eq**) emissions avoided by using stored energy instead of drawing from the grid. CO_2 emitted - CO_2 emitted optimized

Note that if you make more than one submission, only the last one will be considered for evaluation.

3. **Performance metrics must be submitted via FORMS using the "Check Your Results" button:**

   a) **Performance Metric (MAE)** of the prediction for First step Objective 1: **First step of Objective 1: Maximum possible solar generation at each moment on a hourly basis for September 2024.** Note that if you make more than one submission, only the last one will be considered for evaluation. You can use **"the calculator"** to calculate it.

4. **Calculations must be submitted via FORMS using the "Check Your Results" button:**

   a) **Underutilization of the installation (kWh) is requested in Second step of the first objective** The output will consist solely of the total calculation for the period (September 2024), presented as a single figure with four decimal places. You can use **"the calculator"** to calculate it.
   b) **Ra (Self-Consumption Ratio) is requested in First step of the objective 2.** The output will consist solely of the total calculation (%) for the period (September 2024), presented as a single figure with four decimal places. No performance metric is associated with this calculation.
   c) $CO2_{ev}$ is requested in the Objective 3. The output will consist solely of the total calculation (grCO2eq) for the period (September 2024), presented as a single figure with four decimal places. No performance metric is associated with this calculation.

5. **Executive summary and Presentation must be submitted via the "Final submission Form".**

The **executive summary** and the **presentation** to be delivered on the day of the event must be submitted via the "**Final Submission Form**".

- **An executive summary (3-5 pages)** with their inspection strategy and value propositions.
- **Final Presentation.** The presentation should last a maximum of 5 minutes.

Please note: Access to the Google Form is restricted to users with an **IE email account.**

2. FORMS accessible through "Check your Results" button.

You can find the forms accessible through the '**Check Your Results**' button, as well as HERE. **Please note: Access is restricted to those using their IE email account.**

By accessing the same Google Forms where submissions are made, you gain two additional features to help track your progress in the challenge:

- You can see how well you are doing in comparison to other teams by accessing the "Link to Rankings". In this file, you can review the **MAE** metric and the **"Underutilization of the installation", Ra and $CO2_{ev}$ calculations.**

- Additionally, you can check your **MAE and underutilization of the installation (kWh)** metrics by accessing the "Link to MAE Calculator". In this Excel file, you can input two columns with your predicted target metrics and calculate the metrics for September , 2024.

IE Datathon February 2025

A. TECHNICAL EVALUATION CRITERIA:

1. Submissions will be evaluated using the **Mean Absolute Error (MAE)**. The first step of Objective 1 is to determine the **maximum possible solar generation at each moment on an hourly basis for September 2024**.

The model demonstrating superior predictive capability will be the one displaying the smallest values across the MAE metric for the maximum possible solar generation at each moment on an hourly basis for September 2024.

B. BUSINESS EVALUATION CRITERIA:

There are four calculations that will be critical for the business evaluation criteria. A full, detailed explanation of these calculations, the underlying business logic, and their interpretations should be provided in the **executive summary**.

2. **Underutilization of the installation (kWh):** The output will consist solely of **the total calculation** for the period (September 2024)

3. **Ra (Self-Consumption Ratio):** The output will consist solely of the **total calculation (%)** for the period (September 2024).

4. **CO2ev (grCO$_2$ equivalent avoided):** The output will consist solely of the total calculation (grCO2eq) for the period (September 2024),.

Please note: There is no performance metric associated with these calculations. This implies that you cannot compare your predictions/calculations with 'real' data.

C.METRIC'S FORMAT:

Ensure that the metric's format are presented to AT LEAST FOUR DECIMAL PLACES. Submissions that do not adhere to the specified format will not be considered for competition.

D. OUTPUT FILES FORMAT (CSV):

- **Output File 1:** Maximum possible solar generation at each moment on an hourly basis for September 2024. Containing the fields: Datetime and KWH_ENERGIA
- **Output File 2:** Output file that contains the CO2 emissions avoided on an hourly basis for September 2024. Containing the fields: Datetime and CO2_Avoided.

E. CHECK YOUR RESULTS:

- MAE and Underutilization of the installation CALCULATOR (Download a copy).
- LINK TO RANKINGS

E. Evaluation criteria

1. Introduction

Having described all the details of the challenges, including the dataset descriptions and technical delivery requirements in previous sections, we now proceed to explain the evaluation criteria in this section.

2. Evaluation Criteria

The evaluation criteria for this competition encompass both technical (50% of the final grades) and business aspects (50% of the final grades).

- **Technical criteria (50% weight on the final grades**) are meant to assess the general ability of the teams to handle, interpret, and understand the data, build predictive models for the given problem and evaluate their performance according to an objective metric.

- **Business criteria (50% weight of the final grades**) are more general in scope, and they involve the ability to devise what value can be extracted from this data, formulate relevant business questions and try to find answers to them in the data. The quality of the presentation and the ability to communicate clear and powerful ideas on the final pitch will also be part of the business criteria. ***Please be aware that the announcement of finalists will not occur until shortly before the closing ceremony, thus everyone should be prepared to present.***

3. Technical criteria

The models submitted will be evaluated based **not only on the performance metrics but also taking into account additional considerations** described in this section:

- **A. Overall understanding of the problem**
- **B. Exploratory data analysis**
- **C. Feature engineering**
- **D. Predictive models performance**

In addition to this, the code should be clearly structured and results should be interlaced with explanations in jupyter notebooks, Google Colab, etc. The notebooks should be clearly written, and explain the process followed starting from the raw dataset, cleaning and preprocessing, exploratory data analysis, model formulation, hyperparameter tuning (if needed), final metrics and discussion. **The results of the execution must be visible to verify their correctness and quality.**

A. Overall understanding

Ensure that you understand the meaning of each predictor variable in the different datasets: what it means, in which units is it expressed, how is this data registered, at which moment in the time or day, could it contain errors? could it contain outliers ? can we trust the data ? Using common sense, will a given predictor variable be useful to predict our target?

B. Exploratory Data Analysis

Getting acquainted with the datasets is a first necessary step before any modelling on the data takes place. Explore the data distribution, which variables are categorical and which are numerical, do we really understand the meaning of each variable? Are there any correlations among the variables? Are there predictor variables with missing values or outliers? Can we trust the values of the data? Try to formulate hypotheses and understand your datasets before further exploration is conducted. Create good visualizations that help develop your intuition and understand the patterns. If necessary, decide how to handle missing values by either data imputation or removing rows/columns from the dataset. Experiment with various levels of data aggregation to ensure the robustness of your predictions and calculations.

C. Feature engineering

Which features will you use in your predictive model and calculations? Is it legitimate to use all the provided data? Can you imagine how the model will be used in production? Can you enrich your dataset with external information? At a minimum, you will need to combine multiple datasets and/or create new features to train models. Additionally, you should experiment with various aggregation and drill-down techniques to analyze the data effectively. Be creative: anything that you can build on the given data that might have a more direct connection to what you are trying to predict will improve your models performance.

D. Model performance

The models submitted will be evaluated based on the performance metric (MAE) for the target variable, with precision to at least four decimal places. These metrics will assess the predictions for the maximum possible solar generation at each moment, on an hourly basis, for September 2024.

The **superior predictive model** will be determined by the smallest MAE value for the target variable.

Performance metrics and the output files containing predictions and calculations should be submitted via FORMS. **must be submitted via Google FORMS.**

Feel free to try different families of models, adjust their parameters, add regularization, go back to your preprocessing cycle and continue iterating, etc.

Remember that in this edition you can check how well are you doing in comparison with other teams by accessing the "**Link to Rankings**" and also check your MAE metrics by accessing the "**Link to Calculator**" available in Google FORMS.

4. Business criteria (50% weight on the final grades)

In this section, students are expected to demonstrate how they can extract significant value from data by applying model insights to strategic business decisions. A detailed business plan should be articulated, clearly showing the potential environmental impact of their proposed solutions. This plan must be firmly grounded **not only in the maximum possible solar generation and its MAE metric** but also, and more importantly, in the calculated metrics: **Ra (Self-Consumption Ratio) and** $CO2_{ev}$ **(grCO$_2$ equivalent avoided). Clearly explain the results of the optimization that you have achieved.**

Your submission should detail how the implementation of your strategies could enhance the company's commitment to environmental sustainability and create a positive impact on society. This analysis is crucial for demonstrating the real-world applicability and potential effectiveness of your proposed models.

Evaluation Focus:
The evaluation of your business proposal will concentrate on the following areas:

1. **Metric Analysis:** The calculations, explanations, and interpretations of the metrics must be thoroughly detailed. It is essential to provide rich explanations that fully showcase the value and impact of your models through these business metrics. **Clearly explain the results of the optimization that you have achieved:**
   - **Ra (Self-Consumption Ratio)** representing **30% of the final grade.**
   - $CO2_{ev}$ (**grCO_2** equivalent avoided) representing **5% of the final grade.**

The **remaining 15% of the final grade** will include the evaluation of the following aspects of your executive summary and presentation:

2. **Coherence:** The alignment of your business strategy with the technical solutions provided.
3. **Realism:** The feasibility of the proposed business impacts and strategies under real-world conditions.
4. **Executive Report and Presentation Clarity:** the ability to clearly explain the results of the optimization you have achieved, and to communicate your ideas clearly and persuasively.

This competition seeks solutions that are not only technically viable but also use technology to create environmentally sustainable solutions, generating a positive impact on the environment and society.