

---

# DATA DESCRIPTION DOCUMENT: ESG INDICATORS

---

## OVERVIEW

The **Pladifes project (ANR-21-ESRE-0036)** retrieved a dataset of raw data directly sourced from corporates reports. The data is coming from sustainability reports and has been validated and consolidated into an Excel file. It contains more than **13 500 companies** over **2018, 2019 and 2020**, and regroups **15 environmental, social and governance indicators** (when available in the reports).

To date, two datasets are available:

- Corporate ESG raw data, containing raw indicators extracted from company reports.
- Corporate ESG imputed data, an augmented version of the previous database, completed with imputed data.

## CORPORATE ESG RAW DATA

This is the original dataset. It has been transferred from a ceased startup, *Impactfull*. The data is mostly based on companies' sustainability reports and other declarations (annual reports, quarter reports, press publications, *etc.*), and quality checks have been performed to ensure reliability.

To date, the dataset is composed of **26 variables** and more than **40 000 observations**, covering **13 518 corporates** over three year (2018, 2019 and 2020). Companies are among world largest market caps and represent **102 countries** and **154 industries**.

The Pladifes Team is currently investigating on an update of the database.

## CORPORATE ESG IMPUTED DATA

This second dataset is built upon the previous one. Indeed, the previous dataset contains a significant amount of missing values which can limit its usability. We propose an alternative dataset with imputed values to overcome this issue. Please, note that such dataset may not be suited for every application, since data is estimated and cannot be considered as reliable. The used methodology is the one described in the benchmark of this [challenge data](#).

The data in this dataset is of the same format as the previous one, but without any missing value in any column (except in isin and Business Desc.).

The Pladifes Team is currently investigating on implementation of the winning methodology (update planned early 2024).

## DATASET VARIABLES (RAW DATA)

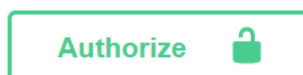
Variable	Description	Unit	Missing values	Type
name_year	Combination of a name and a specific year.	-	-	object
company_id	Company ID.	-	-	object
year	Year (2018, 2019, 2020).	-	-	int64
company_name	The name of the company.	-	-	object
ticker	The stock ticker symbol of the company.	-	-	object
lei_cici	Legal Entity Identifier or financial institution identifier.	-	-	object
isin	International Securities Identification Number.	-	1.5%	object
business_desc	Description of the company's business activity.	-	0.1%	object
region	The geographical region where the company operates.	-	-	object
hq_country	The country where the company's headquarters are located.	-	-	object
primary_industry	The main industry sector of the company.	-	-	object
market_cap	Company size i.e., stock value multiplied by number of shares	\$	-	float64
employees	The number of employees in the company.	-	19%	float64
Revenue	Annual company revenue.	\$	2.4%	float64
scope_1	Company's scope 1 emissions.	TCO2e	72%	float64
scope_2	Company's scope 2 emissions.	TCO2e	72%	float64
scope_3	Company's scope 3 emissions.	TCO2e	74%	float64
waste_production	Annual production of produced waste.	Tons	87%	float64
waste_recycling	Annual quantity of recycled waste.	Tons	92%	float64
water_consumption	Annual quantity of water consumed.	Tons	76%	float64
water_withdrawal	Annual quantity of water withdrawn.	Tons	77%	float64
energy_consumption	Annual electricity consumption.	kWh	85%	float64
hours_of_training	Number of training hours provided to all employees in a year.	Hours	90%	float64
gender_pay_gap	The relative gap between male and female mean salaries.	%	99%	float64
independent_board_members_percentage	Percentage of independent members on the company's board.	%	80%	float64
legal_costs_paid_for_controversies	The total amount of costs linked to controversies.	\$	62%	float64
ceo_compensation	The annual remuneration of the CEO.	\$	94%	float64

## USING THE API

### 1.1. Web interface

To make requests from the interface available on our Pladifes website, follow the following steps:

- Click on the "Authorize" button in green.



- Enter your unique credentials that have been assigned to you.

**Available authorizations**

Scopes are used to grant an application different levels of access to data on behalf of the end user. Each API may declare one or more scopes. API requires the following scopes. Select which ones you want to grant to Swagger UI.

**OAuth2PasswordBearer (OAuth2, password)**

Token URL: token  
Flow: password

username:

password:

Client credentials location:

client\_id:

client\_secret:

- After you have logged in, your session will remain active for 30 minutes. Once this time has elapsed, you will need to log in again to continue using the API. Here is an example request that allows you to download the entire database.

### 1.2. Additional Technical details

The ESG Indicators API is a FastAPI application deployed on Cloud Run after containerization, and it allows to retrieve data stored in a Cloud SQL instance. To access this data from the API, a JSON Web Tokens (JWT) based authentication system is utilized to ensure data security. Please directly reach out to [Mohamed Fahmaoui](#) for more specific information about the data management.

Apart from the web interface, other way of querying the API can be used for more advanced usage.

#### 1.2.1. From Postman

To make a request to a FastAPI API that uses JWT (JSON Web Token) authentication, you'll need to include the JWT token in the header of your HTTP request. Here's how you can do it using the Postman tool:

- Open Postman: Launch Postman on your computer.
- Create a New Request: Click on "New" to create a new request.
- Obtain a JWT Token:

- Choose "POST" as the request type.
- In the "Enter request URL" tab, enter the Authentication URL of the FastAPI Authentication endpoint.
- In the request body, you should typically send your credentials. This could be done in the body as form data, JSON, or other supported formats.
- Send the Request: Click the "Send" button to send the request to the FastAPI Authentication endpoint. If the credentials are correct, you will receive a JWT token in the response.
- Configure the Requested URL: In the "Enter request URL" tab, enter the URL of the FastAPI endpoint you want to call.
- Set Authorization Header: Choose OAuth 2.0 In Authorization and enter the copied token.
- Send the Request: Click the "Send" button to send the request to the FastAPI endpoint with the JWT token included in the header.
- Get the Response: Postman will display the API's response in the response section at the bottom of the interface. You can view the JSON response or any other format returned by the API.

### 1.2.2. From Terminal

To access to the API form CMD terminal, you can use the curl command. Here are the general steps :

- Obtain OAuth 2.0 Access Token: First, you need to obtain an access token from the OAuth 2.0 authorization server. This typically involves making a POST request to the server with your credentials using the following command:

```
curl -X POST -d "grant_type=password&username=YOUR_USERNAME&password=YOUR_PASSWORD&client_id=YOUR_CLIENT_ID&client_secret=YOUR_CLIENT_SECRET" https://api.com/token
```

- Use the Access Token: Once you have the access token, you can include it in your API requests to authenticate yourself.

```
curl -X GET -H "Authorization: Bearer $ACCESS_TOKEN" https://api.com/endpoint
```