Final project part II - Carbon reporting and reduction strategies

Brunno Vanelli

This is the final part of the final project.

The goal of this project is for you to work as a team on solving a real-world problem. Often during development, tasks can come with quite vague requirements and descriptions, even though the final result might be tangible and easy to understand. Your goal as a developer is to come up with a **methodology** to solve the problem, **create a plan**, then execute the plan, **reporting on the progress and the decisions made**.

This project exercises the usage and combination of multiple datasets, with critical thinking and problem-solving skills.

Requirements

- **Form groups of two students** — the same as the previous project — to work independently, you should agree with your partner how each task will be divided, how you are going to collaborate with each other.
- You will use the dataset generated in the previous part of the project in this part.
- Your deliverable will be a **report**. The report will be one Jupyter notebook that will contain the data analysis, as well as the written explanation of the decisions made for analysis. Make sure you include all datasets in the report's folder, so it can be run directly without any additional steps.
- Deliver the report as a `.zip` file on Ilias until the submission deadline. You will also present this report at the end of the semester, in a 5-minute presentation, with 5 minutes for questions. You **do not** need to submit the presentation slides.

The problem to solve

Pretend you work for the IT Department of Albstadt-Sigmaringen University. You have now been assigned a big project for generating a sustainability report.

After the previous task of generating a believable dataset, you are now asked to compute the CO_2 emissions of the university. Data has been provided by entsoe.eu, aggregated by
Electricity Maps. You can use the live tool from app.electricitymap.com to have some ideas about the emissions, since the data is interactive.

The dataset `DE_2024_15_minute.csv` contains aggregated data for every 15 minutes for the year 2024. There are two columns available for emissions: "Carbon intensity gCO_2eq/kWh (direct)" and "Carbon intensity gCO_2eq/kWh (Life cycle)". The direct emissions are the operational emissions and exclude e.g., construction and decommissioning. The life cycle emissions are the ones including construction and decommissioning of power plants and are the ones we are interested in.

Additionally, the columns "Carbon-free energy percentage (CFE%)" and "Renewable energy percentage (RE%)" provide us with context of where this electricity comes from. You can read about the differences in the Electricity Maps API docs.

You can calculate the CO_2 emissions of a point in time by multiplying the consumption, in kWh, by the carbon intensity. You need to account for the different sampling rates between the two datasets.

$$C_{\text{emissions}} = \text{kWh}_{\text{consumption}} \times \text{gCO}_{2\text{eq/kWh}}$$

We want to answer a couple of questions regarding the monitoring of electricity consumption:

- Is the sampling rate for your consumption data (hourly) high enough to accurately measure the CO_2 emissions (given every 15 minutes)? Do you need to implement a system with a higher sampling rate? Calculate the error factor for your dataset using the worst case scenario (here, the error will be also an estimation).
- Calculate the effect of decreasing your sampling rate from hourly to — every 6 hours, every 12 hours, every day and every month. How does the error factor change? Which sampling rate is best for accuracy?
- The University was granted a project for a new Solar Energy installation, report titled *Klimafreundlicher Photovoltaik-Strom für die Hochschule Albstadt-Sigmaringen*, by Baden-Württemberg Ministerium für Finanzen, 22.06.2020. The project is expected to generate 62MWh of electricity per year. Disregarding the emissions for setting up the system and assuming that you can model the electricity production using a tool like PHOTOVOLTAIC GEOGRAPHICAL INFORMATION SYSTEM with hourly data, what are the approximate savings of such a project on CO_2 emissions? **Hint**: You can take the solar data from a different year since it might not be available for 2024.
- Suppose the university wants to acquire a server rack for training of AI models. According to *Data Center Rack Power Costs: A Condensed Analysis*, by Michael Wilson, 11.06.2025, a normal rack specialized for AI training consumes around 60 kW. The rack is expected to train for around 40 hours a week. Come up with a strategy to reduce the CO_2 emissions of such a rack, assuming training can always be time shifted (within the same week). How much CO_2 emissions would be saved in comparison to running overnight (days of the week from 00:00 to 08:00)?
- Download a dataset from the Energy charts for average pricing of electricity in Germany. It provides the average price of electricity in 15-minute windows. Try and correlate, for one small timeslot of 1 week, the average price of electricity and the availability of renewable energy. Can you find a statistical correlation between the variables? **Hint**: You can get this dataset using a JSON endpoint, e.g. Electricity production and spot prices in Germany in week 1 2024
- Can you propose three guidelines, based on the data, for how to reduce the CO_2 emissions? Propose three different strategies, backing your proposal with statistics/visualizations on how those strategies affect the CO_2 emissions. An example strategy would be creating a project to reduce electricity consumption between 02:00 and 6:00 every day (when Carbon intensity is high). I would then evaluate which impact it would have when the consumption could be reduced by 5-10-15% scenarios.

Important to take into consideration:

- Present values in easily interpretable units (percentages, relative changes) rather than large numbers, since humans cannot handle big numbers well. If you present bigger numbers, make sure to convert them to a proper unit, e.g., tons of CO_2.
- A lot of models you will have to approximate, and that is fine. Make sure you **justify** your choices. If in doubt you made the right choice, simulate different scenarios.
- You will be provided with the dataset for CO_2 emissions. If you use other datasets (including your consumption data from Part I), make sure to include them in your final submission.

Presentation

For the presentation, you will quickly present the results of your analysis, focusing on the key points.

The presentation should be 5 minutes long, so focus on what really matters to get your message across. Use statistics and plots to support your decisions. Look at 5-minute-pitch templates for ideas on how to structure it.

Grading

You will be graded on:

- The explanations of the decisions made for each subtask.
- If you have done multiple iterations, the thought process behind each decision: why was each decision made?

- Can you explain the final results in a few sentences?
- Is your notebook well-formatted and easy to read? Is it well-organized? Do the explanations make sense with the code?
- Presentation clarity.
- Being able to answer presentation questions.