
DATA-DRIVEN INSIGHTS FOR SUSTAINABLE DEVELOPMENT GOALS

EPA112A Final Assignment

October-November 2024

INTRODUCTION

The Sustainable Development Goals (SDGs) are a collection of 17 global goals set by the United Nations General Assembly (UNGA). They are part of Resolution 70/1 of the UNGA: "Transforming our World: the 2030 Agenda for Sustainable Development, which is also known as the "2030 Agenda" [1]. The goals are broad and interdependent, yet each has a separate list of targets to achieve. The SDGs cover social and economic development issues such as poverty, health, education, hunger, global warming, sanitation, energy and social justice [2].



The World Bank (<http://www.worldbank.org/>) provides public access to a collection of development indicators, compiled from officially recognized international sources. More specifically, the World Development Indicators (WDI) are large collection of time series from 1960 to 2023, covering topics such as Economic Growth, Poverty, Health, Nutrition, Population, Climate Change, Social Protection and Labour, etc., including national, regional and global estimates [3].

Data analysis helps us to better understand the current state with respect to sustainable developments, the evolution over the past years, to make predictions for the future, and to evaluate effects of different measures.

ASSIGNMENT

For this assignment you will use Python to **perform an original and non-trivial data analysis using the World Development Indicators as a starting point**. While you can use the SDGs as inspiration, you are not required to do so.

Your notebook should follow the structure below:

- **Introduction.** Include a problem statement and your research question(s). You can pick one SDG as a starting point (frame) for your problem.
- **Datasets.** Present the data that you are going to use. Your data must come from at least two sources (one additional source aside from the World Bank data).
- **Preprocessing.** Explain how you clean, format and do any pre-processing work that you find useful, making the data useful for your goals. Implement the steps that you describe in python.
- **Exploration.** Exploratory data analysis – get to know your data.
- **Visualization.** Use multiple types of visualization on the data that make sense for your goal.
- **Machine learning.** Use at least one machine learning technique to make meaningful predictions over (part of) the data.
- **Discussion.** Interpret your results in relation to your research question. Were there any surprises in your research? How reliable do you think your results are? Are there any limitations to your analysis?
- **Conclusion.** What did you learn?

The conceptual contents of the Notebook should be roughly similar to a normal written report of 10-20 pages. Make your Notebook self-explanatory. It should contain text (with references) as well as your actual analysis code and results.

If you want to use other libraries for your analyses or visualizations, feel free to do so. However, only use freely available and well-known libraries. The ones that come standard with Anaconda are fine. If you want to use something else that you first need to install, please check with us first.

Deliverable. Create a zip archive that contains your Notebook and the data you used and upload it on Brightspace. One submission per group.

DATA

WORLD DEVELOPMENT INDICATORS

Relevant indicators drawn from the World Development Indicators, reorganized according to the goals and targets of the Sustainable Development Goals, can be download in .csv format from [here](#) – go to the data and resources tab.

ADDITIONAL DATA ACCESSED DIRECTLY FROM THE WEB

With Python it is possible to read data directly from the web. This way your notebook will always use the most up-to-date data available. In addition to the World Development Indicators imported as .csv files, you are asked to **find one or more data sources that will make your analysis more meaningful**. These additional data sources should be accessed directly from the web. Only use freely available datasets.

One option is to access additional World Bank data using [wbdata](#). Wbdata is a simple python interface to find and request information from the World Bank's various databases, either as a dictionary containing full metadata or as a pandas DataFrame. If you prefer to use data from other websites, you can scrape is using [Beautiful Soup](#).

EXAMPLES

Here are some example questions you might ask **to start** the analysis. Please note that these are just examples and that you are expected to come up with your own questions and analyses.

- What is the average spending on education per country as a proportion of their GDP? Which are the first 5 countries that spend the highest percent of their GDP on education?
- Is there a gender difference in early education completion in low income vs. high income countries?
- The EU has the following goal in the Paris agreement: "At least a 40% domestic reduction in greenhouse gases by 2030 compared to 1990 levels" [4]. How is the EU doing at the moment? If they don't change policy (i.e extrapolation of current trends) where will they end up? Can you classify in good/neutral/bad countries?
- What is the trend per continent with respect to gas emissions? How do countries within the same continent behave? Is it fair to make statements over whole continents?
- Suppose that each country has a % growth or reduction of CO2, where do we end up?
- What is the country with the best trend in renewable energy over the last 10 years?
- Can we predict longevity, natality, or economic growth? How well? And what are good predictors?

GRADING

Some properties of the Notebook and your analyses on which we will grade:

- Use data analysis in Python to answer your research questions. Your code should read, clean and format, process and visualize the data. There should be at least some non-trivial processing involved.

- Demonstrate your skills in Python by using typical Python constructs and using the appropriate data structures.
- Try to think of reusability of your code. How easy would it be to use your same code if we want to try to do a small variations of your analysis? Can we easily adapt/play around with your code?
- How difficult were your analyses?

KNOCK-OUT CRITERIA

- Your notebook should start with a title, your names and student numbers, and your group number.
- Your main file and the zip archive should follow the template:
EPA112A_project_group_groupNumber
- Make sure your Notebook does not generate errors!
- Make sure that your text is readable.
- Make sure that your code is readable:
 - Use sensible variable names,
 - Break up your code into parts, use (fruitful) functions,
 - Use a clean style – check out pep8 style [5],
 - Document your code using comments.
- Use **at least one additional data source to the World Bank data.**
- Use multiple types of visualizations of your results.

Deadline: Thursday Nov 1st, 2024, 23:59.

REFERENCES:

- [1] <https://sustainabledevelopment.un.org/post2015/transformingourworld>
- [2] <http://www.undp.org/content/undp/en/home/sustainable-development-goals.html>
- [3] <https://data.worldbank.org/products/wdi>
- [4] <https://unfccc.int/process-and-meetings/the-paris-agreement/the-paris-agreement>
- [5] <http://pep8.org/>