The link between health status and net greenhouse gas emissions in the EU: a cross-country analysis

1. Introduction

Health condition of the population is one of the most important determinants of the economy, productivity, and prosperity of a population. Among other factors influencing health status are nutrition, individual genetics, and the environment. This project focuses on air quality, specifically net greenhouse gas emission values, as the main proxy for the environmental status of a country. In this project, greenhouse gas emission values and health condition values (measured as healthy life years) of the European nations are analyzed. The following research questions are addressed in the report:

- *What is the relationship between the net greenhouse gas emissions and the health status of EU countries?*
- *Which EU countries exhibit the highest levels of greenhouse gas emissions?*
- *How do net greenhouse gas emission values of the EU countries change over time?*

The results can provide insights into the relationship between greenhouse gas emissions and the health state of the different countries of the EU, contributing to the future policy making and public health strategies.

2. Used Data

As an input for the data pipeline, I used the Eurostat data. Below is a concise snapshot of the two datasets used in the research:

Variable Name	Healthy Life Years (HLY)	Net Greenhouse Gas emissions per capita (GHG)
Domain	Health	Environment
Description	Is an index that measures the number of years a person at birth is expected to live without severe or moderate health issues. The metric integrates mortality data (age-specific death rates/ life expectancy) with health status data from a special survey.	Is an absolute measure that quantifies total greenhouse gas emissions of a country minus the emissions from land use, land use change, and forestry activities.
Measurement	Years	Tonnes of CO2 equivalent per capita
Value-type	Float	Float
Source	Eurostat	Eurostat

The output of my data pipeline (the data pipeline was slightly modified also after the submission deadline) is a merged dataset structured as a panel data, which means they combine time series (data over time) and cross-sectional data (data across different entities, in this case, countries). To comply with the Eurostat's data license obligation, I also state explicitly that some values were modified during the Transform stage of my ETL pipeline. Specifically, I addressed missing values

in the HLY dataset by filling them using the linear interpolation method of neighboring years' values.

The final dataset covers 31 countries with the Healthy Life Years (HLY) data and Net Greenhouse Gas Emissions (GHG) values measured in tonnes per capita for the years 2011-2022.

3. Analysis

The analysis employed a combination of visualization techniques and statistical approaches, specifically focusing on correlation analysis, to delve into how HLY and Net Greenhouse Gas Emissions (GHG) interact over time and across different EU countries.

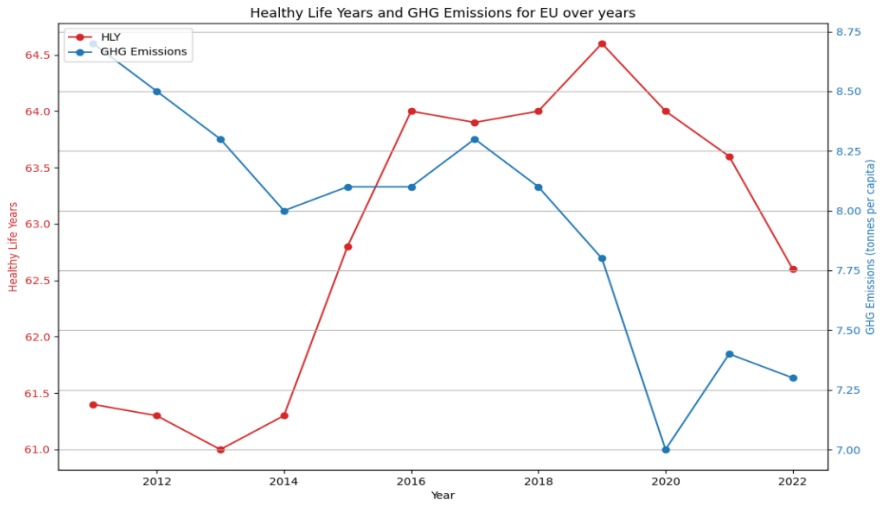
Visual analysis of EU-composite data

The first part of the analysis involved plotting both HLY and GHG for the EU as a whole. For this I use a single "country" measure provided by the Eurostat called “European Union - 27 countries (from 2020)”. This approach allowed to observe a general trend of the two variables across the 27 EU countries.

At the beginning I try to understand the trends of the two variables separately. I notice that HLY does not show a consistent trend and notably declined after 2019, likely due to the COVID-19 pandemic. At the same time net GHG emissions generally decreased over years, which aligns with the EU's sustainability efforts. It is also worth noting that interestingly, in 2020 there was a dramatic drop in emissions. Supposedly, it can be attributed to reduced economic/ production/ transportation activities during the pandemic.

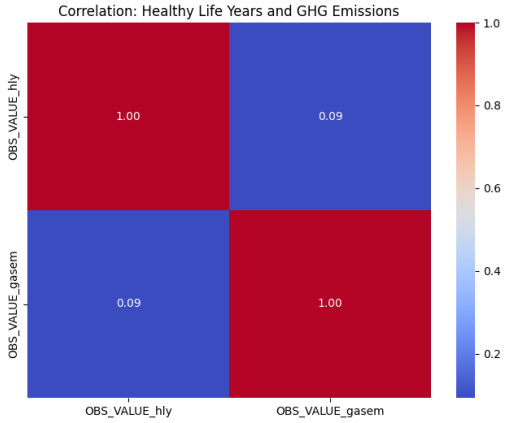
Then I try to visually examine the relationship/association of the two variables together, meaning whether they show co- or counter movements over years. Based on the graph the relationship between HLY and GHG is complex. Certain periods seemed to show negative correlations, such as between 2013-2019, suggesting that improvements in GHG emissions (i.e. lower value of GHG emitted) might align with better health outcomes (more years expected to be lived in a healthy condition). However, other times, such as between 2011-2013 and 2019-2022, surprisingly indicate potential positive correlation, implying that higher net GHG emissions also favor more healthy years expected to be lived by the population. This goes against my initial expectation and a general logical assumption.

Overall, this is suggesting that visual methods alone may not capture the full dynamics at play.



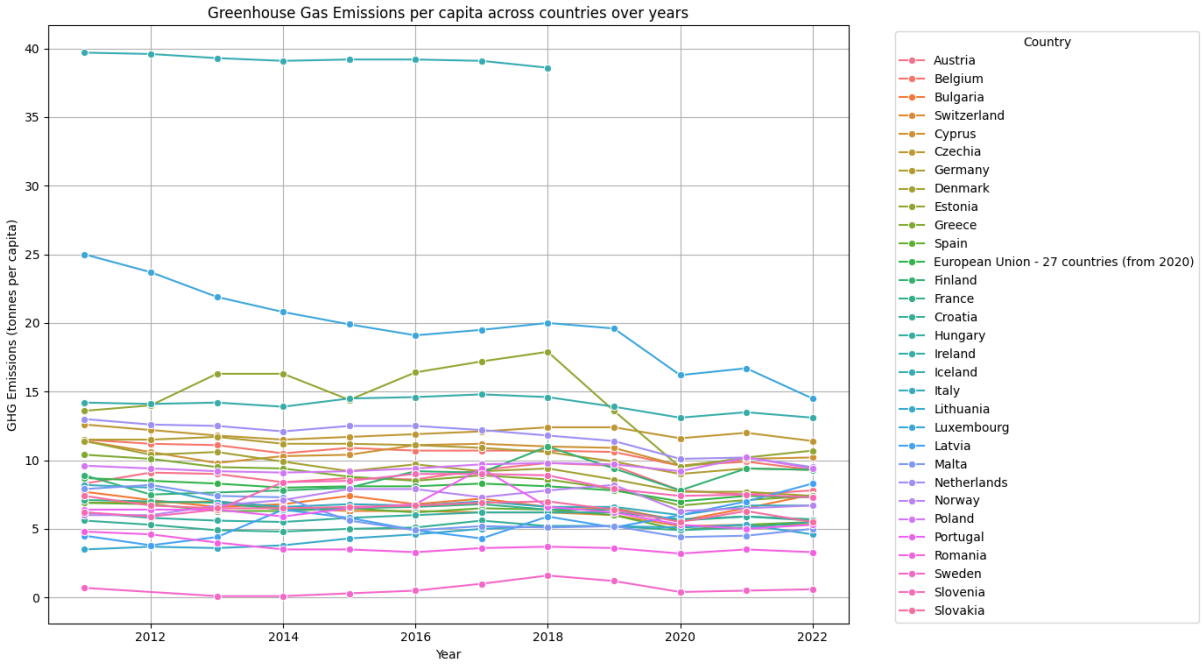
Correlation analysis

Given the mixed visual insights, a more formal statistical correlation analysis was undertaken. The correlation heat map shows a positive but very weak correlation (0.09) between HLY and GHG. Even though the correlation is surprisingly positive, which is an unnatural result given the expected negative correlation (as lower emission levels typically promote better health conditions), I would suggest that this inference is unreliable due to the very low (weak) correlation value. Generally, I would suggest that other factors might be influencing these metrics, therefore my analysis of the relationship between the two variables could not produce meaningful results.



Individual countries' GHG trend over time

I continue my analysis further by plotting the net GHG emissions across individual countries. The graph reveals that most countries emitted similar levels of net GHG per capita: around 3-13 tonnes per capita. Interestingly, Iceland stood out with the highest emissions per capita: almost 40 tonnes of net GHG emissions per capita. This could be attributed to its small population and specific industrial activities, such as aluminum smelting, one of the pillars of the Icelandic economy. Conversely, Sweden demonstrated the lowest emissions among all EU countries, reflecting its strong recycling policies, one of the most environmentally conscious population and extensive use of renewable energy.



4. Conclusion

In this study, I aimed to understand the relationship between public health, measured by Healthy Life Years (HLY), and the environmental impact from net Greenhouse Gas Emissions (GHG) across the European Surprisingly, the analysis showed that the connection between public health and environmental impact is not as direct or straightforward as anticipated. This finding challenges the simplistic assumption that a reduction in GHG emissions would directly lead to improvements in health outcomes proxied by HLY.

A potential reason for the ambiguous findings could be the proxies chosen for the analysis. In particular, HLY is a complex computational index and is therefore prone to biases and errors. Moreover, both of the variables might not perfectly represent the health and environmental domains. So, the choice of the proxy variables represents the main limitation of my research.

The relationship between HLY and GHG may also be influenced by multiple, potentially conflicting factors. For example, these factors might include economic conditions, healthcare access, and lifestyle, with GHG emissions being just one tiny part of this list. It is also worth mentioning that economic crises or most importantly the COVID-19 pandemic can dramatically alter the variables' trends. These complex considerations were out of the scope of my research.

In addressing the other two research questions, I can provide clear answers based on the visual analysis. It was evident that Iceland leads in terms of the highest GHG emissions per capita, likely due to its small population and major industrial activities. Conversely, Sweden is the best performer with the lowest GHG emission values, attributable to its well-established recycling and environmental policies. Overall, the net greenhouse gas emission values of EU countries demonstrate a decreasing trend over time, reflecting the EU's commitment to enhancing sustainability and achieving CO2 neutrality.

In sum, although the initial hypothesis about a direct negative relationship between GHG emissions and HLY was not supported, one can gather some insights from this analysis. They highlight the need for an integrated approach in policymaking that considers the diverse factors affecting public health and environmental impact.