Analyzing the Correlation Between Temperature Changes and greenhouse gas emissions from different sources in Germany

Contents:

1. Introduction

It's important to understand how climate change and gas emission from different sources are related. My data science project looks at how temperature changes and gas emissions specifically CO2 interact in Germany. The goal is to learn things that can help government leaders make plans to reduce the negative impacts of climate change on the environment.

I will be examining how temperature changes and emissions have evolved over the years, focusing specifically on Germany. One of my primary areas of interest is the correlation between temperature fluctuations and different types of emissions within this region. This focus is driven by the need to understand how well these efforts are correlating with changes in temperature patterns and to provide actionable insights that can guide environmental policies in Germany.

In this report, I will explore the following key topics:

- How Have Temperature Changes and Emissions Varied Over the Years in Germany?
- How Are Temperature Changes Correlated with Emissions?

2. Methods & Data Sources

Data Source 1: Food and Agriculture Organization (FAO)

I have selected the FAO Corporate Statistical Database as my primary data source for its extensive range of reliable and comprehensive statistical data. This source is particularly valuable due to its well-maintained records and robust reputation in the field.

Metadata URL 1: FAO Statistics - Greenhouse Gas Emissions
Data URL 1: European GHG Emissions Data (CSV)
Data Type: CSV file containing yearly greenhouse gas emission data for European countries from 1961 to 2023.
License: CC BY-NC-SA 3.0 IGO
The FAOSTAT Emissions Totals domain offers detailed data on greenhouse gas emissions from agrifood systems, tracking emissions such as methane (CH_4), nitrous oxide (N_2O), carbon dioxide (CO_2), and fluorinated gases (F-gases). This data adheres to the IPCC Tier 1 methodology and covers various economic sectors as outlined by the IPCC.

Data Source 2: Food and Agriculture Organization (FAO)

For temperature data, I have again chosen FAO due to their comprehensive environmental statistics.

Metadata URL 2: FAO Statistics - Temperature Change
Data URL 2: European Temperature Data (CSV)
Data Type: CSV file featuring FAOSTAT monthly temperature data for European countries.
License: CC BY-NC-SA 3.0 IGO
The FAOSTAT Temperature Change on Land domain provides statistics on mean surface temperature changes by country, updated annually. This dataset covers the period from 1961 to 2023 and includes monthly, seasonal, and annual mean temperature anomalies.

3. Data Exploration

In the data exploration phase of my project, I focused on organizing and cleaning the datasets effectively to make sure they were ready for analysis. This process involved several specific steps to handle and improve the data related to greenhouse gas emissions and temperature changes across European countries.

Initially, I assessed the datasets' dimensions, noting that the temperature dataset comprised 799 rows and 28 columns, while the emissions dataset contained 1864 rows and 28 columns. My next step was to streamline the datasets by discarding historical data from 1961 to 1999, since I was more interested in the recent data and future projections. This helped make the dataset smaller and easier to manage. This renaming was important to make the data relevant to the upcoming years that are closer and more significant for current decision-making.

Next, I cleaned up the dataset by removing unnecessary columns that were not going to be used in my analysis. For the greenhouse gas emissions data, I removed columns like 'Area Code', 'Area Code (M49)', 'Item Code', 'Element Code', 'Source Code', and 'Source'. For the temperature data, I eliminated 'Area Code', 'Area Code (M49)', 'Months Code', and 'Element Code'. These columns were mostly identifiers and metadata that wouldn't contribute to my analysis focused on trends and relationships. After tidying up the columns, I also made sure to drop any rows that had missing values to maintain the quality and accuracy of my analysis.

4. Initial Inspection

I specifically focused on Germany from the broader dataset of European countries for a more detailed analysis. I concentrated on the period from 2000 to 2023, examining trends in temperature change and CO2 emissions. Among various types of greenhouse gases available in the dataset, I chose to analyze CO2 due to its significant impact on climate change and its relevance in environmental policy discussions.

	Area	Months	Element	Unit	Y2000	Y2001	Y2002	Y2003	Y2004	Y2005	...	Y2014	Y2015	Y2016	Y2017	Y2018	Y2019	Y2020	Y2021	Y2022	Y2023
255	Germany	January	Temperature change	°c	1.662	1.662	1.915	0.462	0.602	2.736	...	3.150	2.951	1.833	-0.984	4.447	1.330	4.355	1.396	3.519	4.237
256	Germany	February	Temperature change	°c	3.787	2.239	4.935	-1.682	2.360	-1.034	...	4.374	0.986	3.350	3.027	-1.688	4.382	5.178	2.032	4.359	3.317
257	Germany	March	Temperature change	°c	1.928	1.106	2.289	2.316	0.839	0.352	...	3.941	2.069	0.938	4.030	-0.791	3.340	2.040	1.696	2.135	2.577
258	Germany	April	Temperature change	°c	2.837	-0.042	0.853	1.076	2.019	1.935	...	3.605	1.230	0.908	0.333	4.945	2.404	3.113	-1.031	0.570	0.327
259	Germany	May	Temperature change	°c	2.887	2.421	2.006	2.363	-0.215	1.068	...	0.776	0.595	1.932	2.399	4.244	-0.758	0.190	-0.848	2.685	1.287

5 rows × 28 columns

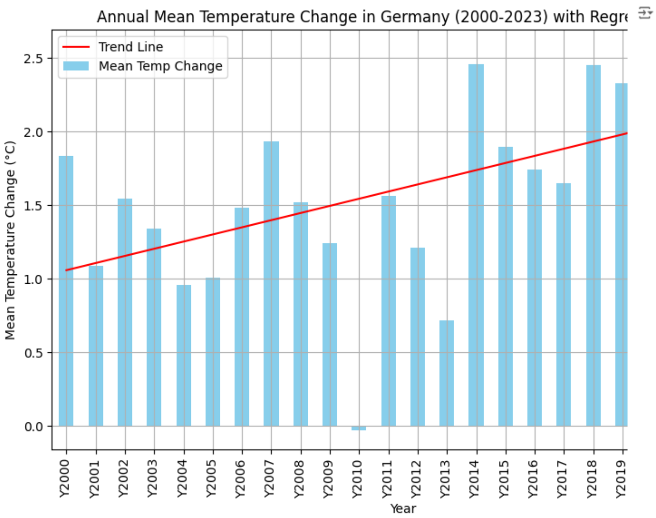
	Area	Item	Element	Unit	Y2000	Y2001	Y2002	Y2003	Y2004	Y2005	...	Y2014	Y2015	Y2016	Y2017	Y2018	Y2019	Y2020	Y2021	Y2022	Y2023
628	Germany	Crop Residues	Emissions (CO2eq) (AR5)	kt	2570.0230	2761.0615	2431.1365	2245.1860	2862.0000	2610.2765	...	2914.7350	2775.8485	2587.4865	2602.9095	2196.5850	2547.6040	2485.4350	2429.9440	2801.3945	2798.5325
633	Germany	Burning - Crop residues	Emissions (CO2eq) (AR5)	kt	145.7856	146.4473	151.0667	155.2342	160.6426	161.2483	...	166.5574	166.5126	159.7724	161.2791	153.0378	156.5856	146.2487	151.2220	152.7636	152.7440
636	Germany	Enteric Fermentation	Emissions (CO2eq) (AR5)	kt	33118.8872	32828.7876	32166.1312	31111.2200	30235.5312	29868.4288	...	29315.9776	29120.3528	28739.3008	28422.9848	27680.5648	27017.2728	26314.2124	25655.2184	28359.0104	26344.3740
643	Germany	Manure Management	Emissions (CO2eq) (AR5)	kt	12387.0721	12340.8117	12255.9816	12088.3917	11768.4839	11916.9371	...	12199.3650	12039.3437	11923.6638	11903.4248	11559.9425	11350.2425	11217.4983	10641.1871	11701.8512	11192.0365
648	Germany	Manure left on Pasture	Emissions (CO2eq) (AR5)	kt	2960.3415	2941.6060	2892.1040	2793.2590	2714.2360	2675.5195	...	2520.8920	2512.4120	2489.2775	2458.6435	2399.7340	2343.2890	2280.1395	2235.0895	2522.2700	2349.9935

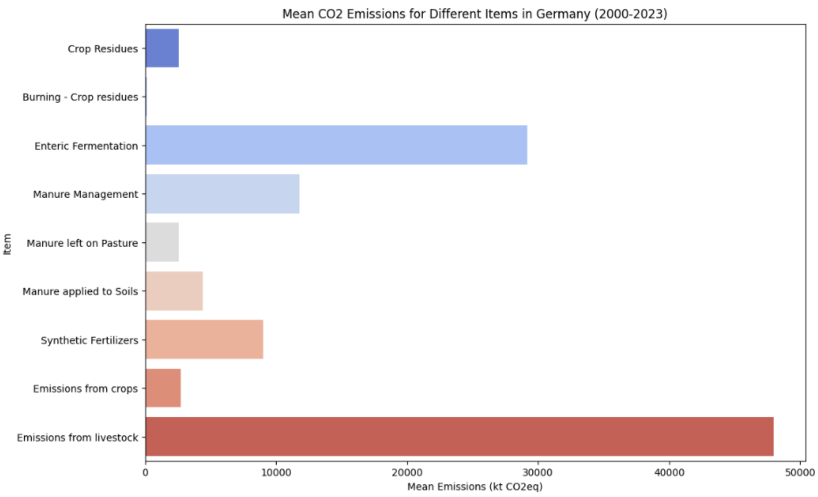
5 rows × 28 columns

5. Results and limitations

- How Have Temperature Changes and Emissions Varied Over the Years in Germany?

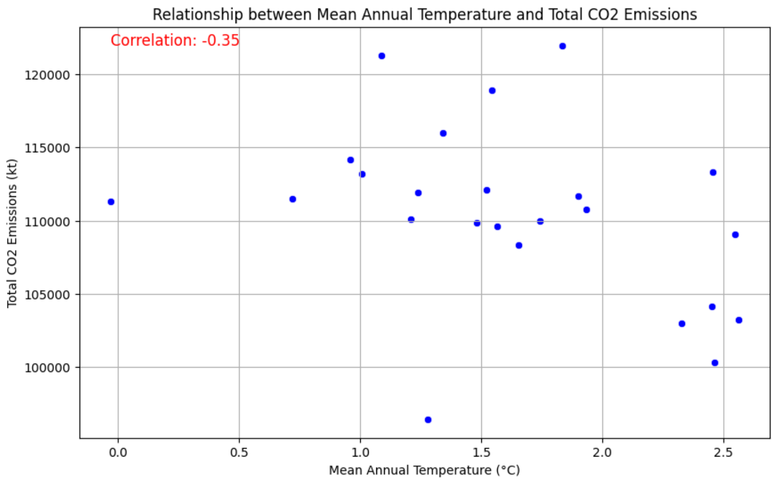
This left graph shows the annual mean temperature change in Germany from 2000 to 2023, represented by blue bars. Each bar indicates the change in mean temperature. The red trend line, derived from a linear regression model, indicates an overall upward trend in mean temperature change over the observed period. The right graph represents the mean CO2 emissions from various sources in Germany between 2000 and 2023. This visualization clearly differentiates the emission levels by source, providing an easy-to-understand comparison of which activities contribute most to Germany's agricultural CO2 emissions over the period. This can aid in identifying key areas for emission reduction efforts.





- How Are Temperature Changes Correlated with Emissions?

The graph illustrates the relationship between mean annual temperature and total CO_2 emissions from the year 2000 onwards. After calculating the yearly averages for temperature and the totals for CO_2 emissions, I merged these two datasets on the year to create a combined dataset. The correlation coefficient calculated was -0.35, indicating a weak negative correlation, suggesting that as the mean temperature slightly decreases, CO_2 emissions tend to increase slightly, and vice versa. This analysis helps in visually understanding the impact of temperatures on emissions over the years.



5.1. Output Data of The Data Pipeline:

The output data consists of two primary SQLite database tables:

emissions_totals.db and temperture_change.db

The Data Structure and Quality of The Result:

- Data Cleaning: I removed irrelevant columns and rows, normalizes data by renaming years, and eliminates rows with missing values to clean the dataset for analysis.
- Standardization: I standardized column names and filters out unnecessary data, making the dataset more manageable and focused on relevant metrics.
- Error Handling: Errors are minimally handled, primarily focusing on encoding issues, but without advanced error recovery or comprehensive logging.
- Data Output: The cleaned and processed data is stored in SQLite databases, allowing for scalable and persistent data management.

Note: I've decided to switch to a new datasets because they provide a better opportunity to understand and solve actual problems. These datasets will help me learn more effectively. I uploaded the data-report.ipynb notebook to project directory. Feel free to take a look at it.