MADE Project Work 7 – Final Analysis Report

Name: Souhardya Chattopadhyay | IDM ID: ow85ymut | Matrikel No.: 23208381

Topic: Impact of Sectoral Energy Consumption on GHG Emissions and Lung Disease Mortality in Germany

1. Introduction

In this data science project, we analyse the relationship between energy consumption, greenhouse emissions, and deaths due to lung diseases such as pneumonia in Germany from 2011-2021. The goal is to understand sector-wise contributions to greenhouse emissions and identify trends for future predictions. The following key questions interests us:

1. Analyse the correlation between greenhouse emissions and total energy consumption in Germany
2. Determine the impact of individual sectors on greenhouse emissions in Germany.
3. Identify trends in GHG emissions in relation to increases or decreases in deaths by pneumonia in Germany.

2. Used Data

The data was extracted from the European data portal (Eurostat). 3 datasets were used representing consumption, greenhouse emissions, and deaths due to lung diseases. Once each dataset was loaded, specific data cleaning tasks were performed and filters to select data for only Germany was implemented on each and then merged into a single data frame, which was used for the final analysis.

2.1. Description and Structure:

Data 1: Energy consumption by sector

The data represents energy consumptions of 4 main sectors: Industrial, Commercial, Household, Transport in TWJ (Tera joules: 10^{12} Joules) from 2011 to 2022. The following columns were present:

- Freq: Cycle of report generation
- nrg_bal: Energy Sector
- siec: Standard International Energy Classification
- unit: Unit of measurement
- geo: Country
- time_period: Year

Data 2: Greenhouse Gas Emissions

The data represents Total GHG Emissions in Kilotonnes from 1990 to 2022. The following columns were present:

- Freq: Cycle of report generation
- airpol: Air pollutant
- src_crf: Source Category
- unit: Unit of measurement
- geo: Country
- time_period: Year

Data 3: Deaths due to lung diseases

The data represents deaths per 100,000 inhabitants from 2011 to 2022. The following columns were present:

- Freq: Cycle of report generation
- unit: Unit of measurement
- sex: Gender of the affected population
- age: Age of the affected population
- icd10: International Classification of Diseases
- geo: Country
- time_period: Year

Eurostat data is typically licensed under open-data licenses, permitting free use with proper attribution. We have complied with these obligations by appropriately attributing the data in our reports and analyses.

3. Analysis

Method: Each of the 3 datasets is in TSV format. One of our first task was to transpose the datasets so that each year which constituted a particular column, get transformed such that each row represents corresponding data of each year. Then we joined the 3 datasets on the year column using outer join to see the overall table. Since 3 datasets had different ranges in year, we saw NaN values. We focussed on the years 2011-2021 for which we had data in all 3 datasets by dropping instances belonging to other years. Finally, we ran correlation analysis and plotted respective graphs to answer our 3 vital questions. Now, let's explore each question and their results in detail:

3.1 Analyse the correlation between GHG emissions and Total energy consumption in Germany.

Result

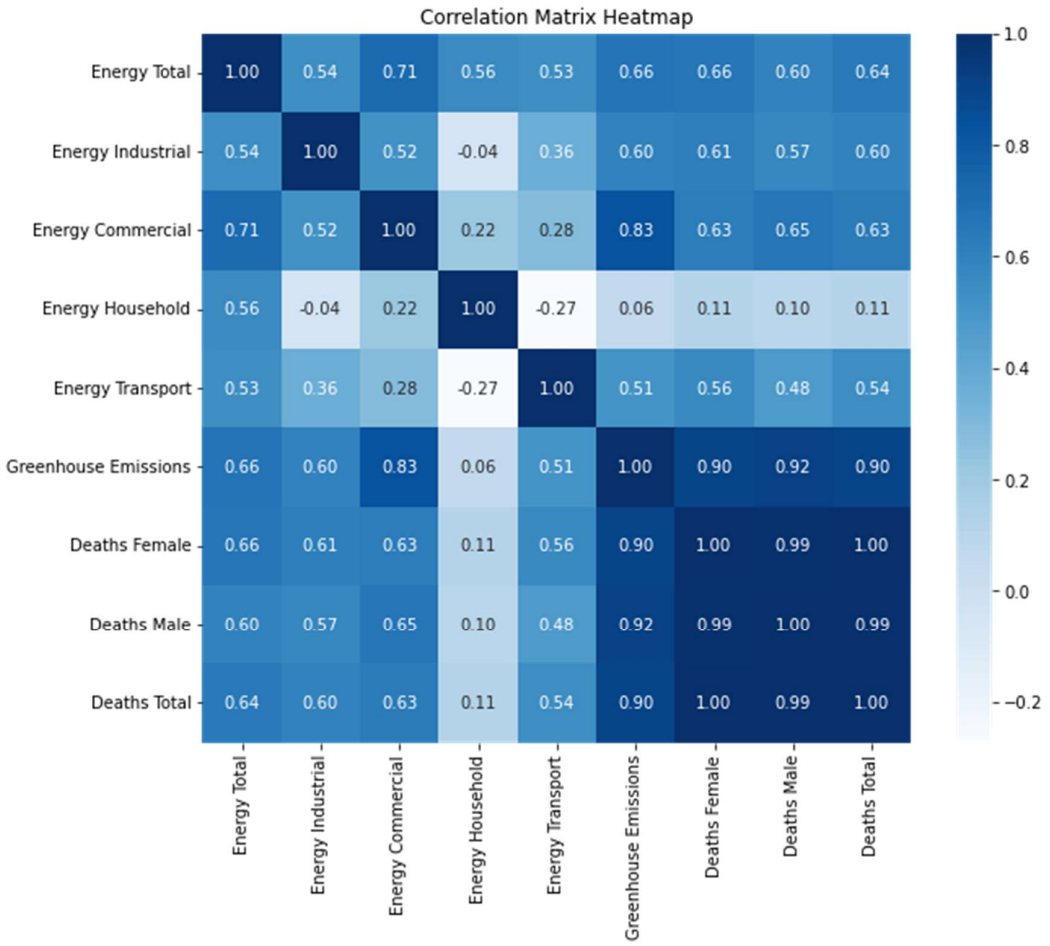

Fig. 1

Interpretations:

- Total energy consumption shows a positive correlation with GHG emissions (r = 0.66), indicating that higher energy consumption leads to increased emissions. This is expected as Germany heavily relies on fossil fuels.
- The industrial sector has a moderate correlation with GHG emissions, reflecting the energy-intensive nature of manufacturing and heavy machinery, which heavily depend on carbon-based energy sources.
- The commercial sector exhibits the strongest positive correlation with GHG emissions (r = 0.83), likely due to significant energy use for heating, cooling, lighting, and electronic devices in commercial buildings, as well as transportation for goods and services.
- The household sector shows no correlation with GHS emissions. Households often adopt energy-saving measures or renewable energy sources, reducing reliance on fossil fuels.
- The transport sector has a moderate correlation with GHG emissions (r = 0.51). Transport, which includes road vehicles, aviation, shipping, and railways, predominantly relies on fossil fuels, contributing to emissions. Efforts to reduce emissions through electric vehicles and public transportation improvements may impact this correlation in the future.

3.2 Determine the impact of individual sectors on greenhouse emissions in Germany.

Result

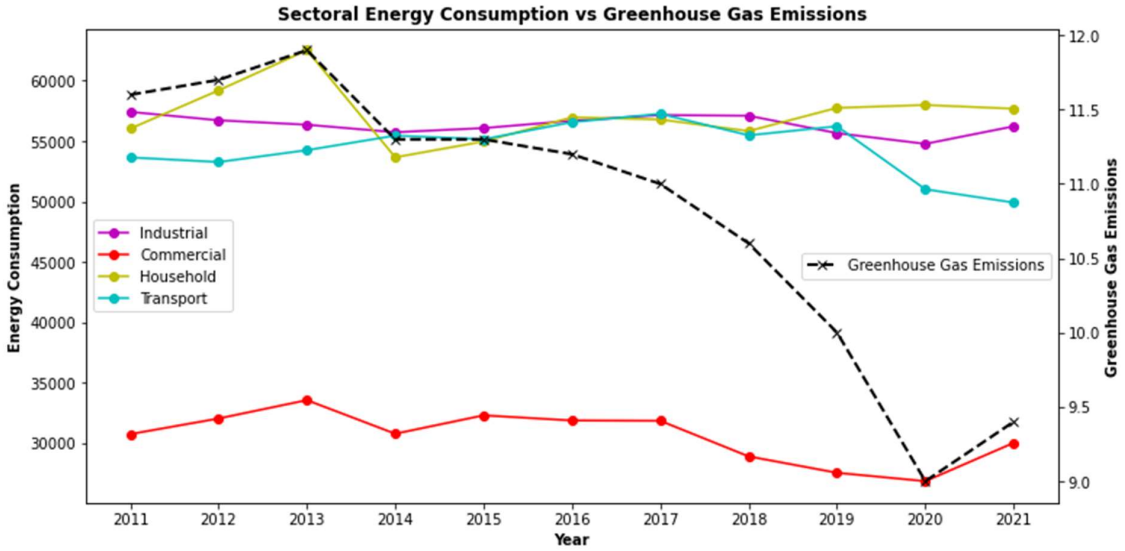

Fig. 2

Interpretations:

- Fig. 1 demonstrates the strongest positive correlation between energy consumption in the commercial sector and GHG emissions, followed by the industrial and transport sectors.

- Fig. 2 highlights the contribution of each sector towards the overall greenhouse gas emissions. We see that Industrial, Household, and Transport contributes the most (in actual numbers) towards the total energy consumption. All 3 sectors range closely between 50,000-60,000 TWh. Industries typically include manufacturing processes and heavy machinery, which often produce significant emissions. Transport includes emissions from road vehicles, aviation, shipping, and railways. Although the Commercial sector shows the strongest positive correlation with GHG emissions, its actual contribution in numbers is half of that of its counterparts.

3.3 Identify trends in greenhouse emissions in relation to increases or decreases in deaths by pneumonia in Germany.

Result

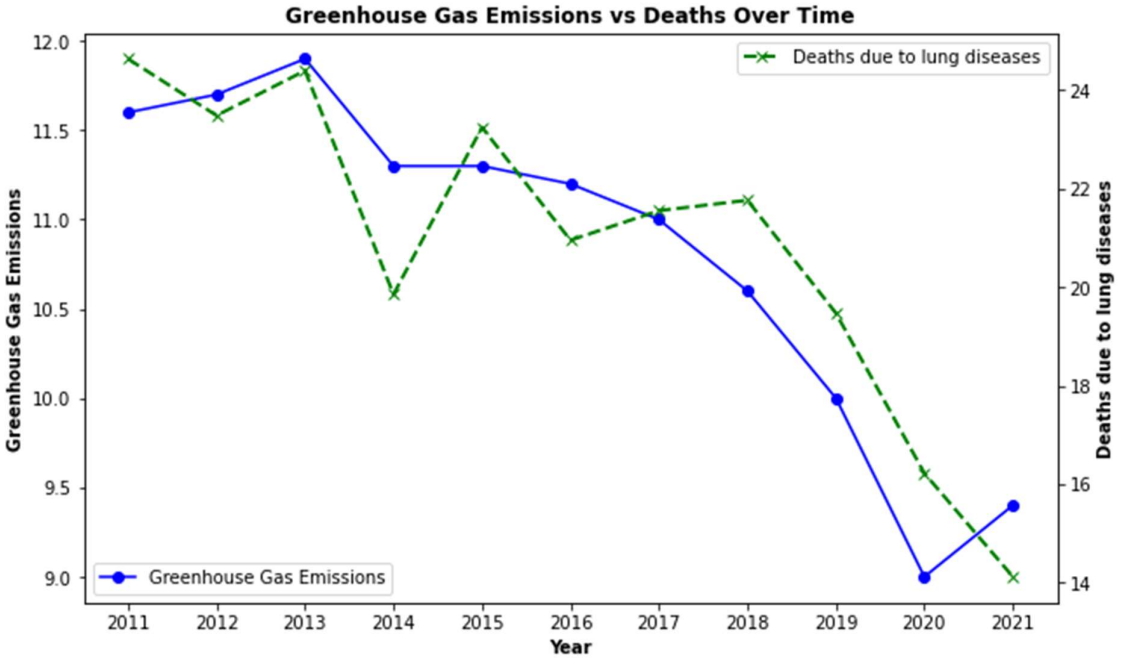

Fig. 3

Interpretations:

- Fig. 1 establishes a strong positive correlation (r ≈ 0.90) between greenhouse gas emissions and deaths due to lung diseases, regardless of gender. The strong correlation between GHG emissions and deaths due to lung diseases highlights the adverse health effects of air pollution, leading to higher mortality rates.
- Fig. 2 indicates a downward slope over the years for both greenhouse gas emissions and deaths almost parallelly. The deaths have reduced to almost half of what it was in the last 10 years from 24 to 14 when the GHG emissions have dropped by 25% of what it was in the same time period from 12 to 9.

4. Conclusions

- While correlation indicates the relationship between energy use and emissions, actual emission contributions reflect the total environmental impact of each sector. Smaller sectors with higher correlation per unit of energy may still have lower total emissions compared to larger sectors. Both metrics are crucial for understanding and managing GHG emissions effectively across sectors. In our case, Commercial sector might have the highest correlation coefficient with respect to GHG of 0.83 but its actual contribution in numbers is half of that of its counterparts, while Household sector might have no correlation with respect to GHG but it contributes the most to the total energy consumption in almost every year.

- Higher greenhouse gas emissions are strongly correlated with increased mortality rates, suggesting a potential health impact associated with environmental pollution. However, our findings suggest that efforts to reduce greenhouse gas emissions may have a significant positive impact on public health, particularly in reducing deaths related to lung diseases.

- Limitation: While correlation indicates a relationship, it doesn't imply causation. Other factors, such as population growth, economic activities, and technological advancements, also influence GHG emissions.