Greenhouse Gas Emissions Footprint Analysis Across Supply Chain Sectors in America

Monika Chavan (23288066)*

Friedrich-Alexander-Universität Erlangen-Nürnberg

January 12, 2025

Abstract

This research conducts a comprehensive evaluation of greenhouse gas (GHG) emission footprints across diverse supply chain sectors in the United States, encompassing manufacturing, transportation, and retail etc. An automated ETL pipeline was meticulously developed to integrate and process extensive datasets with exceptional data quality, enabling a granular examination of emission trends and sector-specific disparities. The study underscores critical variations in carbon outputs among sectors and identifies actionable insights to facilitate targeted decarbonization strategies. Advanced data cleaning, transformation, and analytical techniques rendered the datasets highly functional, while exploratory data analysis illuminated significant patterns and intervention opportunities. This work constitutes a pivotal step toward equipping industries with data-driven tools to minimize environmental impacts and align with national climate objectives. Future iterations aim to incorporate dynamic, real-time emissions tracking and sector-specific normalization frameworks to augment the precision and applicability of the findings.

1 Introduction

Research Question : How do greenhouse gas emissions footprints vary across different supply chain sectors in the U.S.?

To answer this question, we must examine historical emission trends and sector-specific variations in emissions. The goal is to equip industries with the insights needed to develop targeted de-carbonization strategies that align with broader climate goals. This report seeks to fill this gap by leveraging a diverse set of datasets and advanced analytical techniques to explore the GHG emissions within the U.S. supply chain. By examining trends in emissions across different sectors and commodities, this study aims to highlight actionable insights that can help businesses and policymakers target specific areas for improvement. Through a data-driven approach, the report will provide a more nuanced understanding of how different sectors contribute to the overall emissions and the potential pathways for reducing their carbon footprint.

2 Used Data

The data used in this analysis comes from selected datasets spanning over three decades (1990–2022). These datasets focus on different aspects of the U.S. supply chain and offer detailed insights into the GHG emissions produced by different sectors.

1. **Supply Chain Greenhouse Gas Emission Factors v1.3 by NAICS-6 (CSV):** This dataset categorizes emission factors by NAICS-6 codes, allowing for a comparative analysis of emissions across various sectors within the supply chain. Each sector is assigned a unique emission factor, reflecting its GHG impact.

2. **Supply Chain Greenhouse Gas Emission Factors for US Industries and Commodities (CSV):** This dataset focuses on emissions at the commodity level. It provides a granular view of how different materials and products contribute to the carbon footprint across supply chains.

3. **Transportation-Related Greenhouse Gas Emissions (XLSX):** This dataset focuses specifically on the transportation sector and includes detailed emissions data for various transportation modes, highlighting the carbon cost associated with logistics in supply chains.

4. **U.S. Greenhouse Gas Emissions from Domestic Freight Transportation (CSV):** This dataset hones in on the carbon emissions generated by freight transportation, offering detailed temporal data on emissions from trucking, rail, and other freight-related activities.

The data used for this analysis was accessed through open data sources and complies with CC0 and public domain licenses, ensuring free and open use for research and reporting purposes. The datasets are in structured formats (CSV, XLSX) and have been preprocessed to handle missing values, inconsistencies, and unit standardization. The datasets were cleaned and transformed to ensure high-quality, usable data.

3 Approach

This study adopts a structured methodology to analyze greenhouse gas (GHG) emissions across various supply chain sectors, commodities, and time. The approach combines descriptive statistics, correlation analysis, and advanced data visualization to examine emission trends. The specific objectives include:

- Identifying sectors with the highest GHG emissions.
- Analyzing the role of supply chain margins in emission factors.
- Investigating temporal trends in GHG emissions.
- Exploring sector- and commodity-specific contributions to emissions.

An automated ETL (Extraction, Transformation, and Loading) pipeline was employed to streamline the process of extracting, transforming, and loading the data. This pipeline automates data extraction, cleans and transforms it into a structured format, and loads it into a database for subsequent analysis.

3.1 Data Extraction

The extraction phase involves downloading relevant datasets from trusted online sources, stored in CSV and Excel formats. Using Python's requests library, the data is retrieved and stored locally for further processing. The data is then loaded into pandas DataFrames to facilitate efficient analysis.

3.2 Data Transformation and Cleaning

Once the data is extracted, it undergoes several transformation and cleaning steps to ensure its quality like Removing Unnecessary Columns, Handling Missing Data, Reformatting and Standardization.

3.3 Loading Data into the Sink

Following the data cleaning and transformation processes, the processed data is systematically loaded into two primary storage formats for subsequent analysis:

- **SQLite Database (emission.db):** The intermediate data, representing the cleaned but unprocessed data, is stored in the SQLite database `emission.db`.
- **SQLite Database (insights.db):** The final, transformed data is loaded into the `insights.db` database, which serves as the data sink.This database stores the processed data in a structured, with clearly defined schema and standardized naming conventions to support detailed analysis and easy access.

This structured approach ensures that the data is well-organized, easily accessible, and ready for analysis. During the development of this pipeline, one **challenge** was managing the temporary storage of downloaded files. To address this, flag-based approach has been implemented using the `--use-cache` option. This approach ensures that files are only re-downloaded if necessary, optimizing storage space by reusing previously downloaded files when available.

However, A major **limitation** of this current data pipeline is the fact that it is not dynamic with changing data. It does not support updates or changes in the structure of the data, including emission factors or NAICS codes. It also lacks control over versions of datasets, overwriting previous versions without tracking any updates.

Figure 1: Tables in insight.db

4 Analysis

This section outlines the analysis conducted to answer the key research question and to answer this, I have utilized datasets mentioned and applied a systematic methodology. The analysis provided critical insights into the emissions trends across the U.S. supply chain.

4.1 Sector Variability in Greenhouse Gas Emissions

Cement Manufacturing emerges as the most carbon-intensive sector, with emissions of approximately 3.79 kg/2022 USD. These industries are highly energy-intensive, involving processes such as calcination and high-temperature kiln operations, which inherently generate significant CO2 emissions.

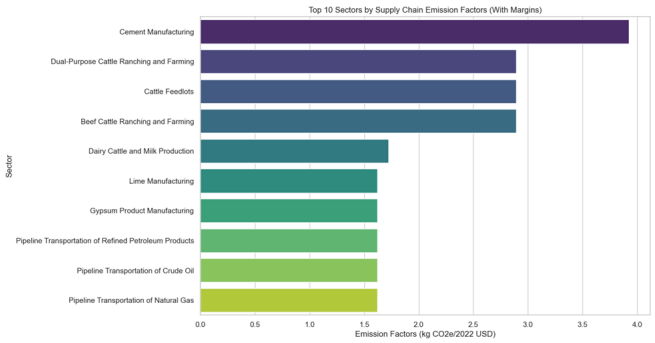

Figure 2: High-Emission Sectors

4.2 Above-Average Emissions

Out of 301 sectors, those surpassing the mean emissions of 0.0102 kg/2022 USD include industries like Adhesive Manufacturing and Basic Organic Chemical Manufacturing, indicating that even less prominent industries can contribute significantly to greenhouse gas emissions.

```
Overall Mean Emissions: 0.010224387893212018
Total Sectors Above Mean Emissions: 301
                                              NAICS Title
1                                 Adhesive Manufacturing
14      All Other Basic Organic Chemical Manufacturing
17     All Other Converted Paper Product Manufacturing
26                             All Other Metal Ore Mining
28   All Other Miscellaneous Chemical Product and P...
```

Figure 3: Mean Emission with Sectors

4.3 Presence of Outliers and Distribution

The emission distribution is right-skewed, with 4,200 outliers indicating that a small number of sectors contribute disproportionately to total emissions.

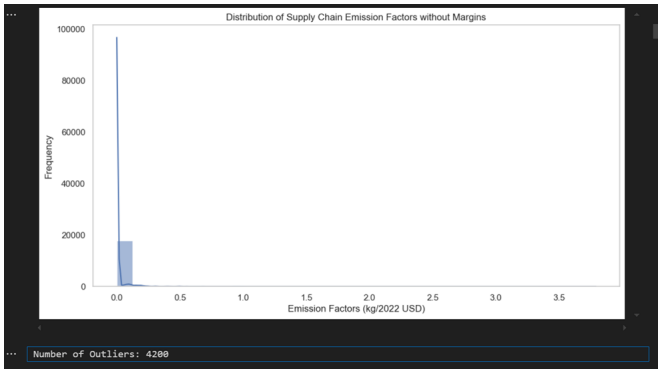

Figure 4: Outliers and Distribution

4.4 Correlation Among Emission Factors

A correlation heatmap indicated a positive relationship between emissions with and without margins, suggesting that emission factors are often consistent across supply chain stages, albeit amplified in margin-intensive sectors.

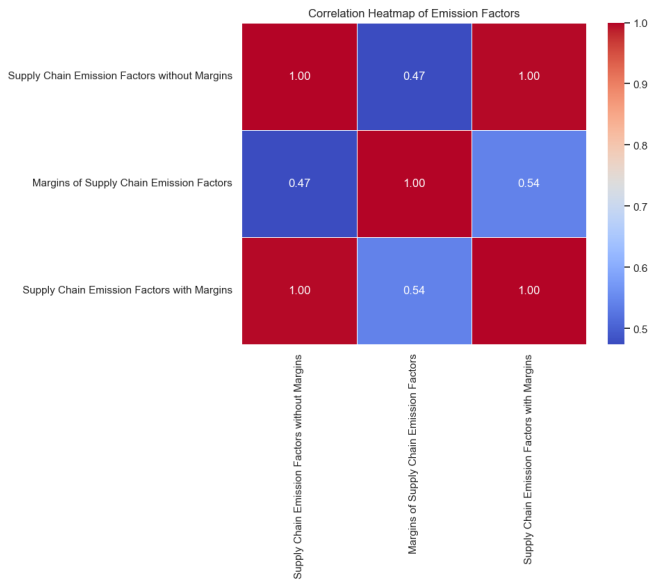

Figure 5: Emissions with and without margins

4.5 Correlation Between Margins and Total Emissions

Strong positive correlations highlight that sectors heavily reliant on intermediary processes are more susceptible to increases in margin-driven emissions.

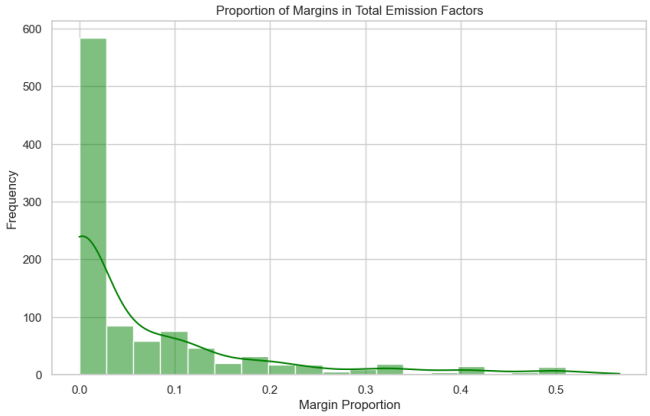

Figure 6: Proportion of Margins in Total Emission Factors

4.6 Temporal Trends and Substance-Level Emissions

From 1990 to the mid-2000s, emissions steadily increased, peaking in 2007, reflecting economic expansion and increased industrial activity. The 2008 financial crisis and the COVID-19 pandemic in 2020 caused notable dips, highlighting the relationship between economic activity and emissions. In observation of emission Factors by substances, The Carbon dioxide is the primary emission contributor, with 0.352 kg/2022 USD, followed by other GHGs (0.013 kg/2022 USD), methane, and nitrous ox-

ide. The dominance of CO2 highlights the need for decarbonization efforts in energy-intensive industries.

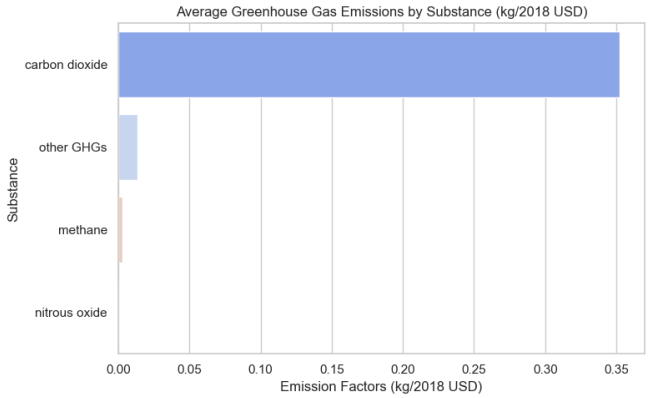

Figure 7: Substance-Level Emissions

4.7 Industry Contributions to Emissions

In high impact category , Utilities, Truck Transportation, and Pipeline Transportation are the leading contributors to greenhouse gas emissions, underscoring the significant carbon footprint of energy production and logistics.

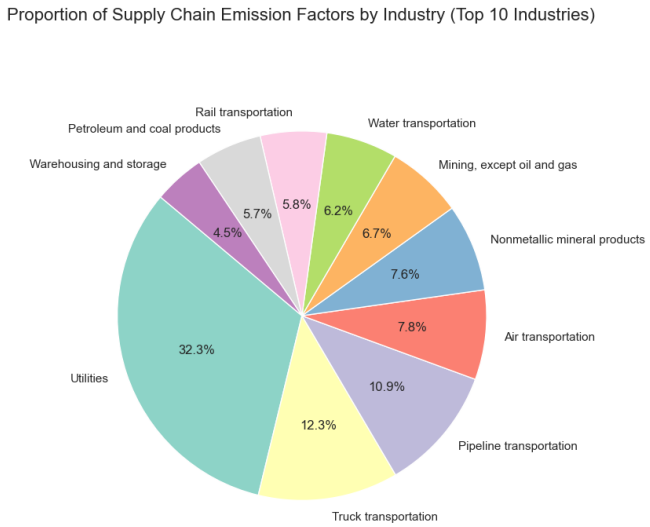

Figure 8: High-Impact Industries Distribution

4.8 Commodity-Level Analysis

A polar plot of GHG emissions across commodities revealed that certain gases (e.g., methane in agriculture, CO2 in transportation) dominate specific sectors, providing actionable insights for targeted mitigation strategies.

5 Conclusion

This analysis reveals substantial variability in greenhouse gas emissions across sectors, driven by industry-specific practices, supply chain complexities, and historical trends. High-emission sectors like Cement Manufacturing and those with significant supply chain margins, such as Utilities and Transportation, contribute disproportionately to total emissions. The temporal shifts in emissions, linked to global economic events like the 2008 financial crisis and COVID-19, indicate the sensitivity of carbon emissions to economic disruptions. However, these reductions are typically short-lived, underlining the importance of long-term, systemic emission reduction strategies. The holistic approaches addressing both direct and indirect emissions, informed by inter-sectoral correlations, hold significant potential for reducing global GHG emissions effectively.

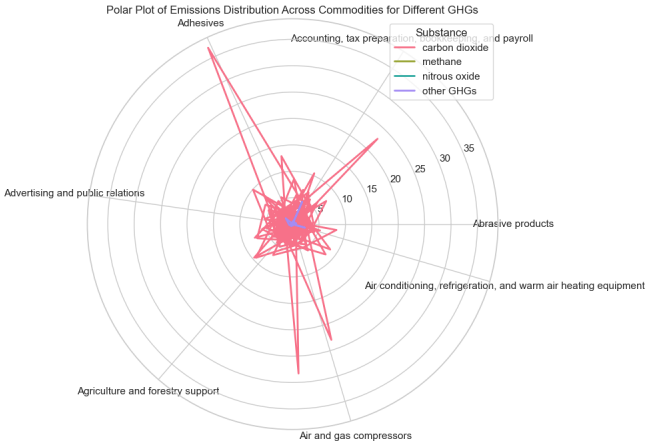

Figure 9: Emission Distribution Across Commodities

6 Future Scope

This study identifies key avenues for advancing research, focusing on emission reduction technologies for carbon-intensive sectors like Cement Manufacturing. Innovations in energy efficiency and carbon capture, alongside improved emission data quality, are critical to effective climate strategies.

References

1. Supply Chain GHG Emission Factors v1.3, *Data.gov*. `https://data.gov`
2. Transportation-Related GHG Emissions, *Bureau of Transportation Statistics*. `https://bts.gov`
3. Greenhouse Gas Emissions in the U.S., *Net0*. `https://net0.com/blog`
4. Carbon Footprints in Supply Chains, *FAO*. `https://openknowledge.fao.org`
5. Efficient Strategies for Supply Chain Carbon Reduction, *arXiv*. `https://arxiv.org/abs/2404.16863`
6. U.S. Supply Chain Carbon Footprint Analysis, *ScienceDirect*. `https://sciencedirect.com`