Investigating Trends in NMVOC Emissions in Ireland: Establishing a Baseline for Climate Action

In the context of Ireland's commitment to climate action, as evidenced by the Climate Action and Low Carbon Development (Amendment) Act 2021, this project focuses on Non-methane Volatile Organic Compounds (NMVOC) emissions. The project aims to establish a baseline by examining historical trends in NMVOC emissions in Ireland prior to the Act's implementation.

NMVOCs are a critical area of study due to their multifaceted impact. They contribute not only to climate change but also pose a threat to human health and agricultural productivity. By analyzing historical data on NMVOC emissions and waste generation, the project seeks to identify potential correlations between these factors. This analysis will provide valuable insights for policymakers as they develop strategies to achieve the ambitious goals outlined in the Act. This project will contribute to Ireland's climate action efforts in the following ways:

- **Establishing a Baseline:** By identifying trends from historical data, the project creates a benchmark for future comparisons, allowing for effective measurement of progress towards NMVOC emission reduction targets.
- **Identifying Key Sources:** Understanding the relationship between waste generation and NMVOC emissions will help pinpoint sectors or activities that require the most significant emission reduction efforts.
- **Optimizing Waste Management Strategies:** Evaluating the balance between waste generation and treatment can inform the development of more sustainable waste management practices, potentially leading to further reductions in NMVOC emissions.

Data sources

GWA01 and GWA02 (Irish Government Website):

Datasource1: Generation of waste (GWA01)	**Datasource2: Treatment of waste (GWA02)**
• Metadata URL: data.gov.ie • Data URL: GWA01 CSV • Data Type: CSV • Published by: Central Statistics Office • Licensed under: Creative Commons Attribution 4.0	• Metadata URL: data.gov.ie • Data URL: GWA02 CSV • Data Type: CSV • Published by: Central Statistics Office • Licensed under: Creative Commons Attribution 4.0

These datasets are sourced from the Irish government's website. The first dataset provides information on the generation of waste categorized by different types of waste from 2004 to 2020. The second dataset details the treatment of waste, including various waste management operations over the same period.

EDGAR v6.1:

Datasource3: NMVOC emissions (NMVOC from EDGARv6.1)

- Metadata URL: https://edgar.jrc.ec.europa.eu/index.php/dataset_ap61
- Data URL: https://jeodpp.jrc.ec.europa.eu/ftp/jrc-opendata/EDGAR/datasets/v61_AP/NMVOC/v61_AP_NMVOC_1970_2018b.zip
- Data Type: xlsx
- Format Extensible Markup Language (XML) file (within a zip archive)
- Condition of data use: Users of the data are obliged to acknowledge the source of the data with reference to the EDGARv6.1 air pollutant website (Metadata URL).

This dataset is part of a larger dataset, EDGARv6.1. The study focuses on NMVOC emission time series (1970-2018) by sector and country and the data provided in an overview table (.xls). Emission country totals are expressed in kt substance/year, and the IPCC 1996 and 2006 codes are used for specification of the sectors.

Key Differences

- GWA01 and GWA02 provide detailed waste treatment data specific to Ireland, downloaded directly from a government website.
- EDGAR v6.1 offers broader emissions data, including NMVOCs, for various countries and sectors, but requires extraction from a zipped archive.

Data usage: The data utilized in this research was sourced under the Creative Commons Attribution 4.0 license. Additionally, we adhered to the condition of data use by acknowledging the source of the data with reference to the EDGARv6.1 air pollutant website (mentioned above). It is important to note that this data was used solely for research purposes and not for any commercial activities.

Data Pipeline

Technology Stack:

- Python: The primary programming language for scripting and data manipulation.
- Libraries: Potential libraries used could include pandas (data manipulation), zipfile (handling zip archives), and sqlite3 (creating and interacting with SQLite databases).

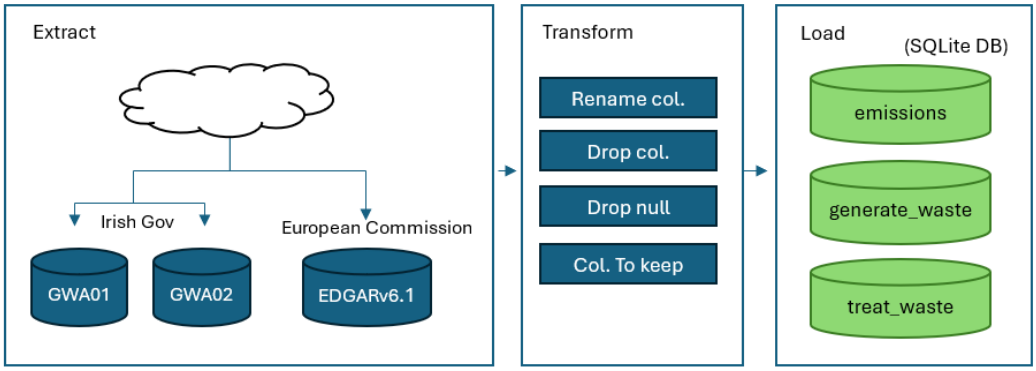

Figure 1 Data Pipeline

Data Processing Steps:

1. **Data Acquisition:**
    - The pipeline initializes with data source information such as URL, name, and output directory.
    - Based on the source format (e.g., zip, CSV), the pipeline employs appropriate methods to retrieve the data.
        - For zipped data, it extracts the relevant file (XLSX in our case) and converts it into a pandas DataFrame.
        - For CSV files, it directly reads them into a pandas DataFrame.
2. **Data Cleaning and Transformation:**
    - The pipeline performs initial cleaning to remove extraneous headers or rows not part of the actual data.
    - Data is then transformed into a structured format using pandas DataFrames. This allows for data exploration and visualization.
    - Unwanted columns are dropped, and relevant columns are renamed for clarity.
    - Rows with missing values are removed to ensure data quality for analysis.
3. **Data Storage:**
    - The preprocessed data is used to create a SQLite database for efficient storage and querying for future use.

While initial data cleaning focuses on generic steps, the pipeline acknowledges the potential for variation across datasets. To handle this, specific filtering of rows and columns is applied after the general cleaning stage and before creating the SQLite database. This approach ensures the final dataset remains consistent as long as the source data format stays the same (no new columns are added).

Result and Limitations

This project's data pipeline delivers structured datasets suitable for further analysis. The data focuses on waste management and NMVOC emissions in Ireland and is stored in SQLite files for efficient querying and analysis.

- **Format:** Structured Data (SQLite)
- **Data Quality:**
  - Accuracy: High. (Data is obtained from reliable sources and is not synthetic.)
  - Completeness: Medium (potential for incompleteness)
  - Consistency: High
  - Timeliness: Date range for our study is between 2004 to 2018. However, it would be more efficient to have more recent data.
  - Relevance: High (directly relevant to NMVOC emissions analysis)
- **Benefits of SQLite:**
  - Lightweight and portable database format. Easy to use and share with collaborators. Efficient querying and analysis using SQL.

Scope of data:

The data might not be entirely comprehensive. Dataset for waste generation and treatment carries information from year 2004 to 2018 which might not be enough to draw solid conclusions about current state of Ireland waste management, but it is enough to give a rough estimation on projection of trends. Though the datasets might seem separate at first, combining them for meaningful analysis is possible. NMVOC emissions from the year 2004 to 2018 show promising trends throughout different sectors and that could be combined with waste generation and treatment to conclude how much the waste sector contributes to NMVOC emissions.

Addressing Limitations:

Obtaining direct download links for the required datasets proved to be a hurdle. Many websites hosting large datasets often mandate user registration. While Selenium automation was considered for scraping the data, the complexity of the webpages involved made it time-prohibitive relative to the potential output. Data range between 2004 to 2018. Having data after 2021 would give insight to effectiveness of Ireland's efforts in reducing greenhouse gases.

Analysis and inference

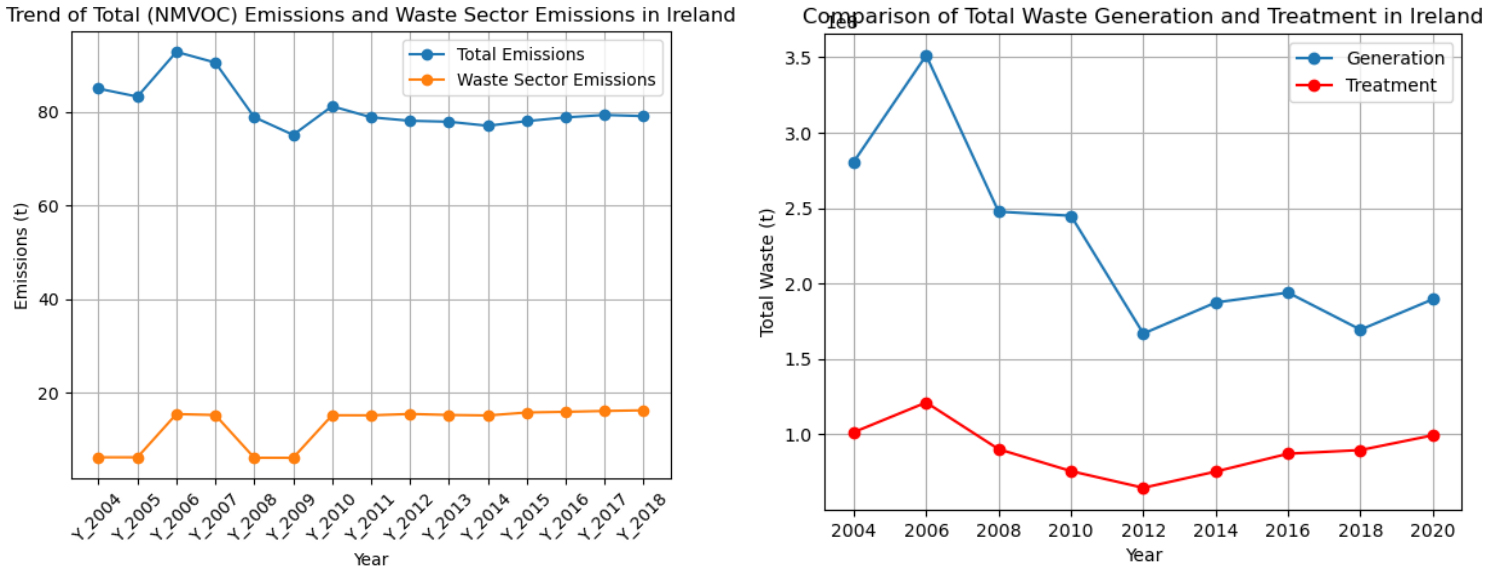

Figure 2 Trends in the data. Plot on the left shows trend between total emissions of NMVOC across multiple sectors and NMVOC emission from waste sector over the years in Ireland. Plot on the right shows total waste generation or treatment over the years in Ireland.

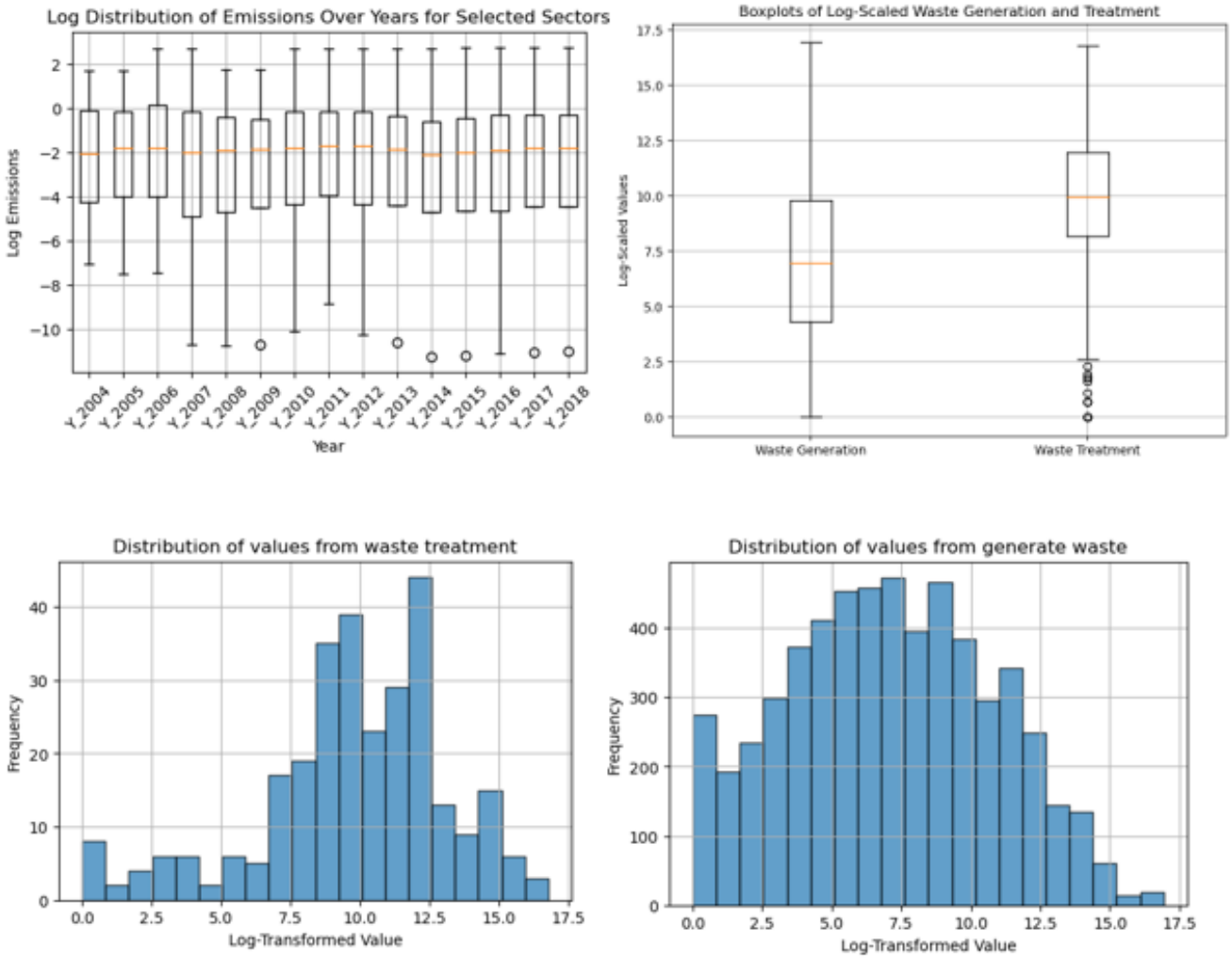

Figure 3 Plots showing distribution of data after preprocessing

The log transformation normalized the data, revealing a normal distribution pattern and facilitating easier interpretation and statistical analysis. Figure 3's bottom plots depict this normalization:

- **Histograms**: These show the frequency distribution of waste treatment and generation values. A normal distribution suggests that most facilities treat or generate waste within a certain range, with fewer extremes. For a more intuitive understanding, a tall bar indicates many facilities generate/treat that much waste, while a short bar represents fewer facilities.

NMVOC Emissions Trends: Figure 2 illustrates the trends observed in NMVOC emissions across various sectors in Ireland. The plot on the left depicts total NMVOC emissions over the years, including a noticeable dip in 2008 and 2009. This dip coincided with the severe economic recession experienced by Ireland during that period. The recovery in emissions thereafter suggests a correlation between economic activity and emissions levels, highlighting the interconnectedness of economic factors with environmental outcomes.

References:

Recession and renewables cut greenhouse emissions in 2009. (n.d.). European Environment Agency. https://www.eea.europa.eu/media/newsreleases/recession-and-renewables-cut-greenhouse

Waste Generation and Treatment: The plot on the right in Figure 2 showcases trends in waste generation and treatment practices from 2004 to 2018. It reveals a decline in waste generation during this period, accompanied by relatively consistent waste treatment practices. This shift signifies potential advancements in waste management efficiency, with more waste being treated despite overall reductions in generation.