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 analyzing 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.

Datasources

GWA01 and GWA02 (Irish Government Website):

Datasource1: Generation of waste (GWA01)

- Metadata URL: https://data.gov.ie/
- Data URL: https://ws.cso.ie/public/api.restful/PxStat.Data.Cube_API.ReadDataset/GWA01/CSV/1.0/en
- Data Type: CSV
- Published by: Central Statistics Office
- Licensed under: Creative Commons Attribution 4.0

Datasource2: Treatment of waste (GWA02)

- Metadata URL: https://data.gov.ie/
- Data URL: https://ws.cso.ie/public/api.restful/PxStat.Data.Cube_API.ReadDataset/GWA02/CSV/1.0/en
- 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 a 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).

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.

Possible data usage: Analyzing how NMVOC emissions have changed over time for different countries and sectors. Comparing NMVOC emissions between different countries or regions. Assessing the impact of different sectors on NMVOC emissions. Evaluating the effectiveness of policies aimed at reducing emissions by comparing data before and after policy implementation. In our case, before the Climate Action and Low Carbon Development (Amendment) Act 2021.

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).

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.

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

Data Considerations:

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 year 2004 to 2018 show promising trends throughout different sectors and that could be combined with waste generation and treatment to conclude how much does the waste sector contribute 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.