ANALYSIS ON THE DATASETS

Dataset-1 Carbon Monitor

a. Dataset URL: https://carbonmonitor.org/

b. Can we access Dataset?: Yes, it is publicly available (Public Domain License)

c. Dataset Topic: The **CarbonEmissionsTool** designed to retrieve and provide detailed carbon emissions data for a specified country, date, and sector, allowing users to gain insights into global or specific carbon emissions trends. The tool interacts with a dataset that includes carbon emissions information from 2019 to August 31, 2024, and provides aggregated results in metric tons of CO2 (MtCO2) per day.

d. Type of Data Available (Chart/Values/...):

- Values (numerical emissions data).

e. Is it live data? Frequency of data updates?

- No, Last data update: September 29th, 2024

f. How can we access the data (Json/Download/View in site)?

- The dataset is available as a CSV file for download.

g. Dataset Pre-processing Need? List any Steps:

1. Change the region names if abbreviated (e.g., "ROW" -> "Rest of the World").
2. Handle missing data, if any.
3. Convert the mixed type of date column into one format(dd/mm/yyyy) for consistency.
4. Convert the timestamp to a human-readable format if needed.

h. Data Summary:

- **Time Period**: The dataset appears to cover multiple years; the starting year is 2019 to 2024 August 31st.
- **Granularity**: Country-level emissions data.
- **Data Structure**:
    - **Rows**: There are multiple entries per country per date, which means a time-series structure.
    - **Columns**: 5 columns in total:
        - country: Country name (e.g., "Brazil").
        - date: Date of recorded data (e.g., "01/01/2019").
        - sector: Economic sector (e.g., "Domestic Aviation").
        - value: Numerical value of emissions.
        - timestamp: Unix timestamp indicating the data point's time.

f. Operations Performed (CRUD): Only reading the data and retrieving the necessary information is done.

Supported Data

The **CarbonEmissionsTool** retrieves data for the following:

- **Time Frame**: 2019 to August 31, 2024.
- **Countries**: Includes 14 specific countries and regions, such as Brazil, China, European Union and UK, India, United States, and more.
- **Sectors**: Includes sectors like Domestic Aviation, Ground Transport, Industry, International Aviation, Power, and Residential.

Key Features

1. **Country-Specific Queries**: Retrieve carbon emissions data for one or more specified countries (e.g., India, Brazil, United States).
2. **Date-Specific Queries**: Filter emissions data by a specific date in 'dd/mm/yyyy' format (e.g., 01/01/2023).
3. **Sector-Specific Queries**: Filter data by sectors such as Transport, Energy, Industry, and more.
4. **Combinations of Filters**: Combine country, date, and sector filters to retrieve precise data.
5. **Error Handling**: The tool provides clear error messages when no data matches the criteria or if the input is invalid (e.g., incorrect date format or unavailable country/sector).
6. **Aggregated Data**: If no filters are provided, the tool returns total carbon emissions across all countries, dates, and sectors.

Types of Queries Supported

The tool supports a variety of queries, categorized by parameters such as country, date, and sector:

1. **Basic Queries**: Provide overall carbon emissions data across all countries, sectors, and dates.
2. **Country-Specific Queries**: Retrieve data for one or more countries, e.g., "What are the total emissions for India?"
3. **Date-Specific Queries**: Filter by a specific date, e.g., "Show carbon emissions for 01/01/2023."
4. **Sector-Specific Queries**: Aggregate data by sectors, e.g., "What are the emissions in the Transport sector?"
5. **Combination of Country, Date, and Sector**: Combine these filters for precise results, e.g., "Show emissions for India in the Transport sector on 01/01/2023."
6. **Handling Missing Data**: The tool provides clear error messages if the requested data is unavailable, or the input is invalid.

Integration with LLM (Llama Stack and Groq API)

The tool is integrated with the **Llama Stack** model through the **Groq API**, which enables seamless interaction between user prompts and the tool:

1. The LLM receives user input and identifies the appropriate parameters for calling the **CarbonEmissionsTool**.
2. The **LLAMA 3.1 70b Instruct Model** invokes the tool based on the user's request, passing the necessary parameters (country, date, sector).
3. The tool processes the request, retrieves the data, and returns results to the LLAMA.
4. The LLAMA/Agent summarizes the results and presents the response to the user.

Example Queries

- "What are the total carbon emissions for Brazil and the United States?"
- "Show carbon emissions data for 01/01/2023."

- "What were the carbon emissions in the Ground Transport sector for India on 15/08/2022?"
- "Provide the total emissions for the Energy sector worldwide."

Error Handling
The tool includes robust error handling for scenarios like:
- Invalid date format.
- Specified country or sector not available in the dataset.
- No data available for the provided criteria.

Some of the Use cases for the CarbonEmissionsTool:

users can ask a variety of questions about carbon emissions data. Here are some types of questions that can be asked, categorized by the different parameters that the tool can handle:

1.Basic Queries (No Specific Filters)
- Example Questions:
    - "What are the total carbon emissions across all countries?"
    - "Show the overall carbon emissions data."

Explanation: These queries don't specify any filters, so the tool will return aggregated data for all countries, all dates, and all sectors.

2. Country-Specific Queries
- Example Questions:
    - "What are the total carbon emissions for India?"
    - "Show the carbon emissions data for Brazil and the United States."
    - "Provide the total emissions data for China."

Explanation: These questions specify one or more countries, and the tool will aggregate emissions data for the specified countries across all dates and sectors.

3. Date-Specific Queries
- Example Questions:
    - "What were the total carbon emissions on 01/01/2023?"
    - "Show the carbon emissions data for 15/08/2022."
    - "Provide the total emissions on 31/12/2021."

Explanation: The tool will filter the emissions data by the specified date, aggregating the results across all countries and sectors.

4. Sector-Specific Queries
- Example Questions:
    - "What are the total carbon emissions in the Transport sector?"
    - "Show the carbon emissions data for the Energy sector."
    - "Provide the total emissions in the agriculture sector."

Explanation: These questions specify a sector, and the tool will aggregate emissions data for that sector across all countries and dates.

5. Country and Date Queries
- Example Questions:
    - "What were the carbon emissions for India on 01/01/2023?"

- "Show the total emissions data for Brazil on 15/08/2022."
- "Provide the carbon emissions for China on 31/12/2021."

Explanation: The tool will filter the emissions data by the specified country and date, aggregating the results across all sectors.

6. Country and Sector Queries

- Example Questions:
  - "What are the total carbon emissions for India in the Transport sector?"
  - "Show the emissions data for the Energy sector in Brazil."
  - "Provide the total carbon emissions for the United States in the Domestic Aviation sector."

Explanation: These queries specify both the country and sector, and the tool will aggregate emissions data for those filters across all dates.

7. Date and Sector Queries

- Example Questions:
  - "What were the carbon emissions in the Transport sector on 01/01/2023?"
  - "Show the emissions data for the Energy sector on 15/08/2022."
  - "Provide the total emissions in the agriculture sector on 31/12/2021."

Explanation: The tool will filter the emissions data by the specified date and sector, aggregating the results across all countries.

8. Country, Date, and Sector Queries

- Example Questions:
  - "What were the carbon emissions for India in the Transport sector on 01/01/2023?"
  - "Show the emissions data for the Energy sector in Brazil on 15/08/2022."
  - "Provide the total carbon emissions for China in the Domestic Aviation sector on 31/12/2021."

Explanation: These queries specify all three parameters—country, date, and sector—allowing the tool to filter the emissions data accordingly.

9. Aggregate Emissions by Sector Across All Countries

- Example Questions:
  - "What are the total carbon emissions in the Domestic Aviation sector across all countries?"
  - "Show the overall emissions in the Energy sector for all countries."
  - "Provide the total emissions in the agriculture sector worldwide."

Explanation: These queries aggregate emissions for a specific sector across all countries and dates.

10. Handling Multiple Countries

- Example Questions:
  - "Compare the carbon emissions for India and China."
  - "What are the total emissions for the United States, Brazil, and Spain?"
  - "Show the emissions data for Russia, Germany, and France in the Transport sector."

Explanation: When multiple countries are specified, the tool will aggregate and provide results for each country, optionally considering the specified date or sector.

11. Complex User Queries based on the data:

- Example Questions:
    - “So, from the emissions of India and Brazil, which country you think should follow strict policies to reduce the emissions.”
    - "What are the total emissions for the United States, Brazil, and Spain? and What suggestions you suggest reducing the emissions?"
    - “What are the total carbon emissions for country India date 01/01/2023 in the sector Ground Transport? What will be the impact in the future if it continues the same way?”
    -

12. Queries Involving Missing Data Handling

Handling the incorrect parameters and if the user's query is outside the supported date range or if the specified country or sector is not in the dataset, inform them of the valid parameters.

- "Show the error message if no data is found for the specified criteria."

Explanation: The tool is equipped to handle cases where no data matches the criteria and will provide an appropriate error message.

These various types of questions allow for flexible querying of the carbon emissions data, covering a range of scenarios that users might encounter. The CarbonEmissionsTool is designed to interpret these queries and provide relevant aggregated results based on the specified criteria.

Setting Up the Tool and Performing Integration Testing

1. Load Files into the Environment

- Extract all the files from ClimateGPT_Model1.zip into your local environment, such as Visual Studio, or a virtual environment like Google Collab.

2. Install Required Dependencies

- Ensure that all necessary libraries and packages are installed in your environment. Use the following command to install dependencies from the requirements.txt file:

pip install -r requirements.txt

3. Install the Llama Stack Client

- Install the Llama Stack Client, which is needed for integrating the tool with the LLM:

pip install llama_stack_client

4. Run the Custom and Carbon Emission Tool Scripts

- Execute the custom tool and carbon emissions tool scripts to initialize the necessary functions:

python custom_tool.py
python carbon_emissions_tool.py

5. Run the End-to-End Testing Script

- After setting up the tools, run the e2e_loop_with_custom_tools.py script to launch the Streamlit interface. This will create the UI for testing the tool integrated with the Llama 3 70b instruct model. Use the following command to start Streamlit:

streamlit run e2e_loop_with_custom_tools.py

6. Streamlit Chatbot Interface for Testing

- Once Streamlit launches, a chatbot UI will appear. This interface allows you to interact with the tool and perform unit testing. You can test different queries related to carbon emissions using the chatbot, where the tool is integrated with the Llama 3 70b instruct model.