

Analyzing the Impact of US Industrial Sectors on Climate Change and Economic Growth (2010-2020)

Main Question

How do different industrial sectors in the United States contribute to both economic growth and climate change, and what are the relationships between sectoral GDP contribution and their corresponding environmental impacts from 2010 to 2020?

Description

The relationship between industrial growth and climate change is a critical challenge facing the United States, because industries are both essential drivers of economic prosperity and major contributors to greenhouse gas emissions. This project analyzes the dual impact of major US industrial sectors on economic growth and climate change, using statistical analysis of industry-specific GDP data and environmental impact data. The results can give insights into which industries provide the best balance of economic benefits versus environmental costs, helping policymakers and business leaders make informed decisions about industrial development and environmental regulation.

Data Sources

Datasource 1: Greenhouse gas emissions by sector

Metadata URL: [Our World in Data](#)

Data URL: [Greenhouse gas emissions by sector](#)

Data Type: CSV

Description: The dataset breaks down annual greenhouse gas emissions by major economic sectors globally, covering energy use, industry, agriculture, and other sectors from 2010-2020, helping understand which parts of the economy contribute most to climate change.

Licensing and Obligation: As cited on the [website](#) “*Licenses: All visualizations, data, and articles produced by Our World in Data are open access under the Creative Commons BY license. You have permission to use, distribute, and reproduce these in any medium, provided the source and authors are credited*”

Datasource 2: Interactive Access to Industry Economic Accounts Data

Metadata URL: [BEA](#)

Data URL: [Industry Economic Accounts Data](#)

Data Type: CSV

Description: This dataset tracks the economic performance and GDP contributions of different U.S. industries over time, providing detailed metrics on industry value added, output, and growth rates.

Licensing and Obligation: As cited on the [website](#) “*BEA makes its statistics available to the public in many formats. This page provides links to BEA's major data tools and datasets.*”

Data Pipeline

To create the ETL process for 2 datasets separate ETL pipelines are implemented in python.

Emissions ETL:

Extraction:

- Retrieves CSV data from OWID's emissions dataset
- Uses requests to fetch data from a [predefined URL](#)
- Handles potential request exceptions

Transformation:

- Converts raw CSV data into a pandas DataFrame
- Performs basic data cleaning:
 - Converts 'Year' column to numeric
 - Sorts data by year and entity
- Manages potential transformation errors

Load:

- Saves transformed data to a CSV file
- Creates a 'data' directory if not exists
- Displays dataset information and first few rows

Emissions Data Pipeline

- Orchestrates the entire ETL workflow
- Coordinates extraction, transformation, and loading
- Provides error handling for the complete process

BEA ETL:

Extraction:

- Retrieves GDP by industry data from BEA API using developer API key in JSON format
- Configurable with API key
- Handles API request exceptions

Transformation:

- Converts raw JSON data into a pandas DataFrame
- Performs advanced data transformations:
 - Filters specific industries
 - Pivots data by year
 - Adds custom line numbers and descriptions
 - Rounds numeric values
- Provides hierarchical data representation

Load:

- Saves transformed data to a CSV file
- Creates a 'data' directory if not exists
- Displays dataset information and first few rows

BEA Data Pipeline

- Orchestrates the entire ETL workflow
- Coordinates extraction, transformation, and loading
- Provides error handling for the complete process

Problems Encountered, Solutions, and Error Handling

- **Schema Inconsistency** between Data Sources was found as for this we applied pivoting on years column in BEA datasource
- **Inconsistent Columns Names** between both dataset, e.g in one name of industrial sector is fishing but in the other it is mentioned fishing and agriculture, so we ensured consistent column names across both datasets.
- **Error Handling** becomes crucial in managing these unpredictable data scenarios. Both pipelines implement error management strategies, but with notable differences. The BEA GDP pipeline demonstrates a more comprehensive approach by providing detailed error logging, including error type, specific message, and full traceback information. This allows for more precise debugging and understanding of potential failure points during data extraction and transformation.
- **Retry Mechanisms** were implemented. Added automatic retry for transient network errors and Implement exponential backoff for API requests

Results and Limitations

Data Output: The output of the data pipeline consists of transformed datasets in CSV format. For the Greenhouse Gas Emission Dataset, the output includes a separate CSV file for the United States of America from the years 2010 to 2020. Similarly, for the World GDP dataset, the output is a CSV file containing GDP data for industries for the USA from 2010 to 2020.

Data Structure, Quality and Format: The output datasets maintain a tabular data schema with data types for each attribute. The quality of the output is ensured for further analysis through various data cleaning and transformation steps, including filtering for relevant countries and years, dropping unnecessary columns, and renaming columns for clarity. considering where the data sources come from, the reliability also can be ensured. Moreover, CSV (Comma-Separated Values) is chosen as final format because it is widely used, easily accessible, and compatible with various data analysis tools like pandas in Python.

Reflection and Potential Issues: While the data pipeline successfully processed the data and no missing data was found for the given country and years, potential outliers may still appear during the analysis