

Data Report

Question

How have carbon emissions changed over the last 20 years across different countries?

Data Sources

World Bank Open Data

- **Why Chosen:** The World Bank Open Data provides comprehensive historical data on carbon emissions for various countries, making it an excellent source for analyzing long-term trends.
- **Data Origin:** This data is sourced directly from the World Bank's official database.
- **Data Content:** It includes metrics on CO2 emissions (in metric tons per capita) for numerous countries from the year 1960 onwards.
- **Data Structure and Quality:** The data is structured in a CSV format with columns for country name, country code, and yearly CO2 emissions. The quality of the data is high, with consistent formatting and minimal missing values.
- **Licenses:** The World Bank data is available under an open data license that allows for free use, provided proper attribution is given. We plan to include the necessary attribution in any publications or presentations that use this data.

Global Carbon Atlas

- **Why Chosen:** The Global Carbon Atlas provides additional data on CO2 emissions from fossil fuels, cement production, and land-use changes, which complements the World Bank data.
- **Data Origin:** This data is sourced from the Global Carbon Project.
- **Data Content:** It includes detailed emissions data for a wide range of countries, focusing on various sources of CO2 emissions.
- **Data Structure and Quality:** The data is provided in a CSV format, with each row representing a country-year pair and the corresponding CO2 emissions. The quality is generally good, though some inconsistencies and missing values are present.
- **Licenses:** The Global Carbon Atlas data is also available under an open data license, allowing for free use with proper attribution. We will ensure to comply with the license terms by providing appropriate attribution in our work.

Data Pipeline

Overview

Our data pipeline automates the process of downloading, cleaning, and transforming data from the World Bank and Global Carbon Atlas. The pipeline is implemented using Python, leveraging libraries such as pandas, requests, and selenium.

Technologies Used

- **Python:** The primary language for scripting and data manipulation.
- **Pandas:** For data cleaning and transformation.
- **Requests:** For downloading data from the World Bank.

Transformation and Cleaning Steps

1. **World Bank Data:**
 - Downloaded the data as a zip file and extracted the relevant CSV.
 - Removed unnecessary columns and reshaped the data to a long format with columns for year, country, and CO2 emissions.
 - Converted the year column to integer and handled any missing CO2 emissions data.
2. **Global Carbon Atlas Data:**
 - Automated the download using Selenium to navigate and click the download link on the website.
 - Cleaned the data by removing the first row (containing column names), renaming columns appropriately, and reshaping the data to have year, country, and CO2 emissions columns.
 - Converted the year column to integer and ensured all CO2 emissions values are numeric.

Challenges and Solutions

- **Challenge:** The World Bank data URL initially led to malformed data issues.
 - **Solution:** Updated the code to correctly handle zip file extraction and read the CSV with the appropriate encoding.
- **Challenge:** Automating the download from the Global Carbon Atlas required handling a dynamic webpage.
 - **Solution:** Used Selenium for browser automation, set up a profile for handling file downloads, and ensured the script waits for the download to complete.
- **Challenge:** Handling large datasets efficiently and dealing with missing or inconsistent data.
 - **Solution:** Utilized pandas' efficient data manipulation capabilities and optimized the data processing steps to minimize memory usage and processing time and implemented data cleaning steps such as removing or imputing missing values and standardizing data formats across datasets.

Error Handling

The pipeline includes error handling for:

- Failed data downloads (retries or logs the issue).
- Missing or malformed data (logs and skips problematic rows).
- Non-numeric data in numeric fields (coerces to NaN and drops these rows).

Result and Limitations

Output Data

- The output data includes cleaned and structured CSV files for both World Bank and Global Carbon Atlas data.
- **Data Structure:** Each output CSV has columns for year, country, and CO2 emissions.
- **Data Quality:** High quality, with missing and inconsistent data handled appropriately.

Data Format

- **CSV:** Chosen for its simplicity and ease of use in data analysis tools and programming languages.

Reflection and Potential Issues

- **Data Completeness:** Some countries may have missing data for certain years, which could affect the analysis.
- **Consistency:** Differences in data collection methods and definitions between sources could introduce inconsistencies.
- **Future Work:** We plan to continuously monitor the data quality and update our pipeline to handle any new issues or data formats.

Conclusion

Our automated data pipeline successfully downloads, cleans, and transforms carbon emissions data from multiple sources. The resulting datasets are ready for analysis to answer our primary research question. Despite some challenges, our approach ensures high data quality and robustness against common data issues.

Pipeline Structure

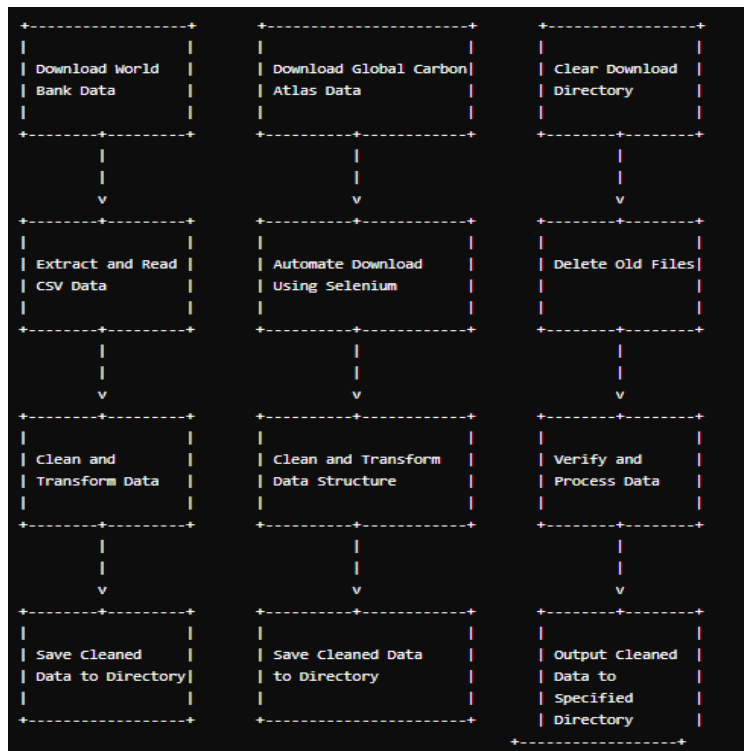


Figure 1: Overview of the Data Pipeline Structure

Data Sources and Sample Data

World Bank Data

Data Source,"World Development Indicators",
Last Updated Date,"2024-05-30",
Country Name,"Country Code","Indicator Name","Indicator Code","1960","1961","1962","1963","1964","1965","1966","1967","1968","1969","1970","1971"
Aruba,"ABW","CO2 emissions (kt)","EN.ATM.CO2E.KT",,,,,,,,,,,,,,
Africa Eastern and Southern,"AFE","CO2 emissions (kt)","EN.ATM.CO2E.KT",,,,,,,,,,,,3046.
Afghanistan,"AFG","CO2 emissions (kt)","EN.ATM.CO2E.KT",,,,,,,,,,,2046.87",1941.37",,
Africa Western and Central,"AFW","CO2 emissions (kt)","EN.ATM.CO2E.KT",,,,,,,,,,,,9719.
Angola,"AGO","CO2 emissions (kt)","EN.ATM.CO2E.KT",,,,,,,,,,,,6564.2","6674.5","6877.3"
Albania,"ALB","CO2 emissions (kt)","EN.ATM.CO2E.KT",,,,,,,,,,,,6060.5","4119.6","2239.3"
Andorra,"AND","CO2 emissions (kt)","EN.ATM.CO2E.KT",,,,,,,,,,,,406.704","406.704","406.
Arab World,"ARB","CO2 emissions (kt)","EN.ATM.CO2E.KT",,,,,,,,,,,,632472.87","634670.8"
United Arab Emirates,"ARE","CO2 emissions (kt)","EN.ATM.CO2E.KT",,,,,,,,,,,,55210.4","6."
Argentina,"ARG","CO2 emissions (kt)","EN.ATM.CO2E.KT",,,,,,,,,,,,100313.9","105906.6",,
Armenia,"ARM","CO2 emissions (kt)","EN.ATM.CO2E.KT",,,,,,,,,,,,19845.2","20689.6",,
American Samoa,"ASM","CO2 emissions (kt)","EN.ATM.CO2E.KT",,,,,,,,,,,,218.2",221.
Antigua and Barbuda,"ATG","CO2 emissions (kt)","EN.ATM.CO2E.KT",,,,,,,,,,,,263437.5","264709.7",,
Australia,"AUS","CO2 emissions (kt)","EN.ATM.CO2E.KT",,,,,,,,,,,,58275.7","62423.2",57668.
Austria,"AUT","CO2 emissions (kt)","EN.ATM.CO2E.KT",,,,,,,,,,,,53483.7",52059.4",63.
Burundi,"BDI","CO2 emissions (kt)","EN.ATM.CO2E.KT",,,,,,,,,,,,187.6",229.6",201.2",21.
Belgium,"BEL","CO2 emissions (kt)","EN.ATM.CO2E.KT",,,,,,,,,,,,109312.6","113960.1",141.
Benin,"BEN","CO2 emissions (kt)","EN.ATM.CO2E.KT",,,,,,,,,,,,330.11",273.51",282.61",,
Burkina Faso,"BFA","CO2 emissions (kt)","EN.ATM.CO2E.KT",,,,,,,,,,,,594.1",599.6",610.
Bangladesh,"BGD","CO2 emissions (kt)","EN.ATM.CO2E.KT",,,,,,,,,,,,11523.73",10830.76",,
Bulgaria,"BGR","CO2 emissions (kt)","EN.ATM.CO2E.KT",,,,,,,,,,,,73605.7",58286.5",5536.

Figure 2: Initial Sample Data from World Bank

Global Carbon Atlas Data

Territorial emissions in MtCO ₂																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																
,Afghanistan,Albania,Algeria,Andorra,Angola,Anguilla,Antigua and Barbuda,Argentina,Armenia,Aruba,Australia,Austria,A																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																
1960,0.41388,2.0222,6.1512,,0.54895,,0.03664,48.7645,3.1539,0.61856,87.4443,30.7845,18.9845,0.41037,0.57525,2.2119,0.172																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																

Figure 3: Initial Sample Data from Global Carbon Atlas