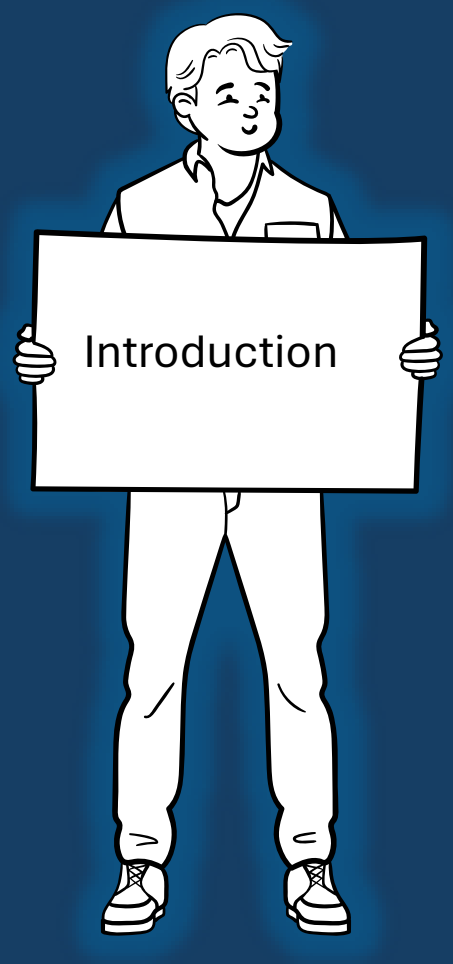
Analysis of Carbon Emissions Over the Past 20 Years

Presented by: Eldrin Sanctis
Matrikel no: 23079169



To analyze carbon emissions trends over the past 20 years using data from the World Bank and Global Carbon Atlas to understand patterns and effectiveness of global efforts to reduce emissions

Data Source

World Bank Open Data

- Provides historical CO2 emissions data from 1960 onwards.
- High quality data, minimal missing values.

Importance of Dataset

- Detailed emissions data focusing on fossil fuels, cement production and land-use changes
- Inconsistencies and missing values

Data Pipeline

Python for scripting and data manipulation

Pandas for data cleaning and transformation.

Requests for downloading data

Selenium for automating data download from dynamic web pages.

Data Cleaning and Transformation

World Bank Data

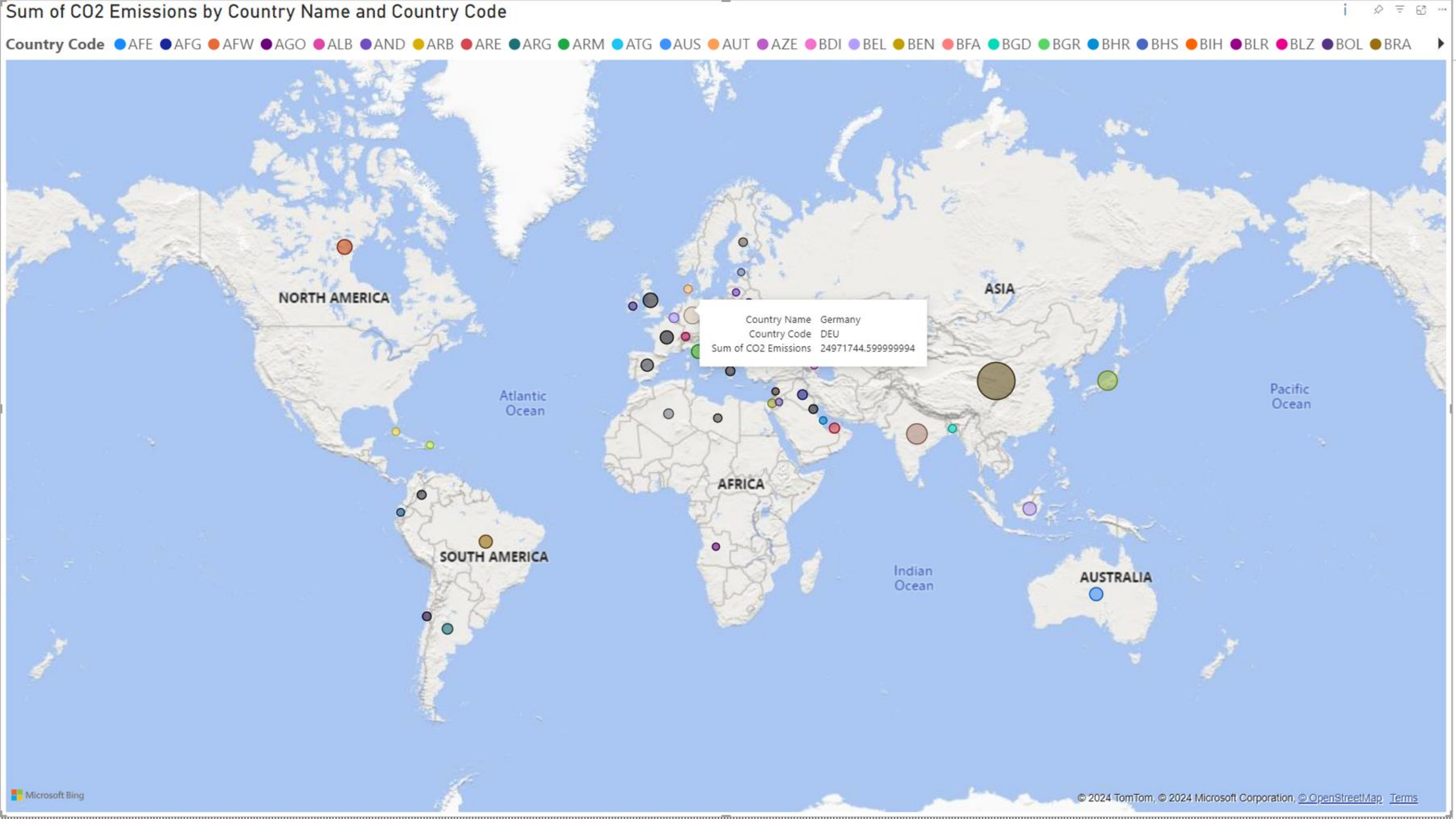
- Downloaded as a zip file, extracted to CSV.
- Removed unnecessary columns, reshaped data.
- Handled missing CO2 emissions data

Global Carbon Atlas Data

- Automated download with Selenium
- Removed first row, renamed columns.
- Ensured CO2 emissions values were numeric.

Data Visualization

Sum of CO2 Emissions by Country Name and Country Code



Challenges and Solutions

Challenges

- Malformed data issues
- Dynamic webpage handling for data download.
- Large dataset with missing/inconsiste nt data.

Solutions

- Correct handling of zip extraction.
- Browser automation with selenium.
- Efficient data manipulation with pandas.

Conclusion:

Successful automation of data pipeline.

High-quality, ready to analyze datasets.

Future Work:

Continuous monitoring and updating of data pipeline.

Addressing data completeness and consistency issues