Data Report: EV Adoption and Climate Change

Question

This project investigates whether the increased adoption of electric vehicles (EVs) contributes to measurable reductions in greenhouse gas (GHG) emissions, helping to mitigate climate change.

Main Question: Does the growth in EV adoption correlate with a reduction in CO_2 emissions in the Americas?

This analysis integrates data on EV adoption trends and GHG emissions to provide insights into the environmental impact of EVs.

Data Sources

1. Electric Vehicle Data

- **Source**: International Energy Agency (IEA)
- **Dataset URL**: Global EV Data Explorer
- **Description**: Historical data on EV sales, cumulative stock, and charging infrastructure by year and country.
- **License**: Open for non-commercial use. Obligations fulfilled by citing the source and linking to the license.

2. Emissions Data

- **Source**: World Bank Open Data
- **Dataset URL**: CO_2 Emissions
- **Description**: Annual CO_2 emissions per capita by year and country. Useful for analyzing trends and regional differences.
- **License**: Open under World Bank's terms. Attribution and license compliance fulfilled.

Data Summary Table

Dataset Name	Source	Key Variables	Coverage	Format
EV Adoption Data	IEA	Year, Country, EV Sales	Americas (Global)	CSV
CO_2 Emissions Data	World Bank	Year, Country, CO_2 Emissions	Americas (Global)	CSV

Data Pipeline

Overview

The pipeline automates the collection, cleaning, and transformation of raw data into a structured SQLite database.

Data Report: EV Adoption and Climate Change

Pipeline Steps

Data Loading

- Loaded raw CSV files into Pandas DataFrames.
- Ensured data paths were dynamically handled for local and hosted environments.

Data Cleaning

- Renamed columns for consistency.
- Addressed missing values using forward-fill.
- Skipped malformed rows using `on_bad_lines='skip'`.

Data Transformation

- Integrated datasets by matching `year` and `country`.
- Added calculated columns for EV market share, where applicable.

Data Storage

- Stored cleaned and transformed data in an SQLite database for querying.

Figures and Tables

Figure 1: Pipeline Workflow

This diagram shows the data flow from raw data collection to final storage in an SQLite database.

Table 1: Example of Cleaned EV Data (Top 5 Rows)

Year	Country	EV Sales	Cumulative EV Stock
2015	United States	120,000	450,000
2016	Canada	15,000	55,000
2017	Brazil	5,000	18,000
2018	Mexico	8,000	30,000
2019	Argentina	2,000	7,000

Table 2: Example of Cleaned Emissions Data (Top 5 Rows)

Year	Country	CO_2 Emissions (tons per capita)
2015	United States	16.5
2016	Canada	15.8
2017	Brazil	2.1

Year	Country	CO_2 Emissions (tons per capita)
2018	Mexico	3.9
2019	Argentina	3.5

Results and Limitations

Output Data

- **Format**: SQLite database with two tables (`ev_data`, `emissions_data`).
- **Structure**:
  - `ev_data`: Contains EV adoption metrics by year and country.
  - `emissions_data`: Contains CO_2 emissions metrics by year and country.

Quality Improvements

- Addressed missing data with imputation techniques.
- Unified column names for consistency.

Limitations

1. **Data Gaps**: Some countries lack complete emissions or EV data.
2. **Bias**: EV data is more comprehensive for developed nations, skewing trends.
3. **External Factors**: Reductions in emissions may also be influenced by renewable energy adoption or economic downturns.

Conclusion

The data pipeline successfully prepares high-quality datasets for analyzing the relationship between EV adoption and CO_2 emissions. While the pipeline ensures data quality and consistency, future work will address gaps and biases in the datasets.