

Analyzing the correlation between carbon emission and the adoption of renewable energy across different countries.

Main Question

How do carbon emissions correlate with renewable energy consumption in different nations?

Data Sources

Datasource 1: Global Carbon Atlas

- **Reason for choosing the data:** The Global Carbon Atlas provides one of the most comprehensive datasets available on carbon emissions and is updated regularly.
- **Source:** Global Carbon Project
- **Content:** Data is structured annually and includes total CO2 emissions, fossil fuel combustion, cement production, and gas flaring, emissions per capita.
- **Metadata URL:** <https://globalcarbonatlas.org/emissions/carbon-emissions/>
- **Data URL:** https://zenodo.org/records/10562476/files/GCB2023v43_MtCO2_flat.csv
- **Data Type:** CSV
- **License:** CC BY 4.0
- **License Compliance:** The license allows for any use of the data for non commercial purposes provided that the citation is given.
- **Citation** Andrew, R. M., & Peters, G. P. (2024). The Global Carbon Project's fossil CO2 emissions dataset (2023v43) [Data set]. Zenodo.

Datasource 2: The World Bank Open Data

- **Why Chosen:** This dataset includes statistics on renewable energy consumption by country.
- **Source:** World Bank
- **Content:** The dataset concludes the information of access to electricity, renewable energy generation, renewable energy consumption in TJ, and etc.
- **Metadata URL:** <https://databank.worldbank.org/source/sustainable-energy-for-all#>
- **Data URL:** <https://www.kaggle.com/datasets/hungenliao/global-sustainable-energy-from-1996-to-2016?select=Sustainable.csv>
- **Data Type:** CSV
- **License:** CC BY 4.0
- **License Compliance:** The license allows for any use of the data for non commercial purposes provided that citation is given.
- **Citation** World Bank. (2024). Sustainable Energy for All. DataBank. Retrieved from <https://databank.worldbank.org/source/sustainable-energy-for-all#>

Data Structure

Displaying the carbon emission and renewable energy consumption data from source

```
In [13]: import pandas as pd
```

```
import opendatasets as od
from sqlalchemy import create_engine

carbon_emission = pd.read_csv('https://zenodo.org/records/10562476/files/GCB2023v4
carbon_emission.head(5)
```

Out[13]:

	Country	ISO 3166- 1 alpha- 3	UN M49	Year	Total	Coal	Oil	Gas	Cement	Flaring	Other	C
0	Afghanistan	AFG	4	1750	NaN	NaN	NaN	NaN	0.0	NaN	NaN	
1	Afghanistan	AFG	4	1751	NaN	NaN	NaN	NaN	0.0	NaN	NaN	
2	Afghanistan	AFG	4	1752	NaN	NaN	NaN	NaN	0.0	NaN	NaN	
3	Afghanistan	AFG	4	1753	NaN	NaN	NaN	NaN	0.0	NaN	NaN	
4	Afghanistan	AFG	4	1754	NaN	NaN	NaN	NaN	0.0	NaN	NaN	

```
In [6]: dataset = 'https://www.kaggle.com/datasets/hungenliao/global-sustainable-energy-fr
od.download(dataset)

renewable_energy = pd.read_csv('global-sustainable-energy-from-1996-to-2016/Sustai
renewable_energy.head(5)
```

Skipping, found downloaded files in ".\global-sustainable-energy-from-1996-to-2016"
(use force=True to force download)

Out[6]:

	Country Name	Country Code	Time	Time Code	Access to Clean Fuels and Technologies for cooking (% of total population) [2.1_ACCESS.CFT.TOT]	Access to electricity population [1.2_ACCESS.ELECTRI
0	Afghanistan	AFG	1996.0	YR1996	..	
1	Afghanistan	AFG	1997.0	YR1997	..	
2	Afghanistan	AFG	1998.0	YR1998	..	
3	Afghanistan	AFG	1999.0	YR1999	..	
4	Afghanistan	AFG	2000.0	YR2000		8.8

DATA PIPELINE

Description

Technology Stack and Steps

- **Extraction:** Pandas and opendatasets libraries have been used to automate the data extraction process.
- **Transform:** Pandas has been used for transformation and data cleaning.
- **Load:** SQLite database has been used to save the data after the transformation.

Data transformation and cleaning:

- **Dropping columns:** Data irrelevant to the analysis have been removed from both the datasets

- **Handling Invalid Data:** Invalid data is removed from the dataset. Columns have been checked for correct data type and invalid data inputs.
- **Missing Data:** Missing values have been filled with median of neighboring values.
- **Renaming Columns:** Some columns have been renamed for easier data exploration.

Problems Encountered and Solutions

- **Problem:** missing values in important columns.
 - **Solution:** Missing data has been dropped or imputation technique has been used to fill them up.

Error Handling

- **Error Handling:**
 - Implemented try-except blocks in Python to catch and log errors during data ingestion.

Results and Limitations

Output Data

- **Format:** The transformed data is saved in the tables: emission and renewable in a sqlite database named ProjectDatabase.
- **Reason:** SQLite is chosen as it is lightweight and it reads and writes directly to disk files. The serverless architecture and portability makes sqlite a good choice for lightweight projects.

```
In [17]: # display output tables
engine = create_engine('sqlite:///../data/ProjectDatabase.db')
carbon = pd.read_sql_table('emissions', engine)
carbon.head()
```

```
Out[17]:
```

	Country	Year	Total	Coal	Oil	Gas	Cement	Flaring	Oth
0	Afghanistan	1996	1.370104	0.007328	0.985616	0.307776	0.047400	0.021984	(
1	Afghanistan	1997	1.304152	0.003664	0.948976	0.282128	0.047400	0.021984	(
2	Afghanistan	1998	1.278504	0.003664	0.941648	0.263808	0.047400	0.021984	(
3	Afghanistan	1999	1.091640	0.003664	0.776768	0.241824	0.047400	0.021984	(
4	Afghanistan	2000	1.047128	0.003664	0.787760	0.223504	0.010216	0.021984	(

Data Quality

- **Data Integrity:** While the data is reliable as it comes from dedicated sources, discrepancies in reporting methodologies across countries can introduce variations in data accuracy.

Critical Reflection and Potential Issues

- **Data Availability:** As the datasets do not cover a wide range of years it might result in analytical gaps.