Methods of Advanced Data Engineering (MADE) Analysis Report

Berk Tunc, 23072225

July 4, 2024

1. Introduction

This report explores the relationship between CO2 emissions and population growth of the countries. Analysing the relationship between CO2 emissions and population growth of countries is crucial as it helps identify the impact of population dynamics on environmental sustainability, guiding policy decisions for emission reductions. The **main question** addressed is: Which countries have the lowest amount of increasing CO2 emissions compared to their population growth?

2. Data Sources

To answer the question, two data sources have been selected for this project: both of them are from The World Bank. First one provides CO2 emissions(in kilotons) data for all countries in the world and the second one shows the total population numbers for all countries in the world.

Data source 1: CO2 Emissions Dataset

Metadata URL: https://data.worldbank.org/indicator/EN.ATM.CO2E.KT

Data URL: https://api.worldbank.org/v2/en/indicator/EN.ATM.CO2E.KT?downloadformat=csv

Description: The World Bank dataset captures the annual CO2 emissions for various countries over the period from 1990 to 2020. Each row represents a specific country and each column from 1990 to 2020 contains the CO2 emissions value for that year. Each entry in the dataset includes the country name, country code, year and the corresponding CO2 emission value.

License and Obligations: Attribution-NonCommercial 4.0 International (CC BY-NC 4.0). Link to the license: https://creativecommons.org/licenses/by-nc/4.0/

Changes made to dataset: Columns and rows that were not required for analysis are filtered.

Data source 2: The Total Population Dataset

Metadata URL: https://data.worldbank.org/indicator/SP.POP.TOTL

Data URL: https://api.worldbank.org/v2/en/indicator/SP.POP.TOTL?downloadformat=csv

Description: The World Bank dataset captures the annual population figures for various countries over the period from 1990 to 2020. Each row represents a specific country, and each column from 1990 to 2020 contains the population value for that year. Each entry in the dataset includes the country name, country code, year and the corresponding population value.

License and Obligations: Creative Commons Attribution 4.0 (CC-BY 4.0). Link to the license: https://datacatalog.worldbank.org/public-licenses#cc-by

Changes made to dataset: Columns and rows that were not required for analysis are filtered.

3. Analysis

First rows of datasets which are output of my data pipeline:

(Actually, datasets have values for the years 1990-2020. This is a simple demonstration.)

CO2 Emissions Data:

	Country Name	Country Code	1990	1991	1992	1993	1994	1995
0	Africa Eastern and Southern	AFE	304614.720181	300135.943094	296834.691723	305267.779142	314596.287085	329784.557767
1	Afghanistan	AFG	2046.870000	1941.370000	1525.470000	1527.890000	1493.590000	1457.690000
2	Africa Western and Central	AFW	97190.345000	110559.933800	121628.003000	114995.869000	106198.491000	116198.614000

Population Data:

	Country Name	Country Code	1990	1991	1992	1993	1994	1995	1996
0	Aruba	ABW	65712.0	67864.0	70192.0	72360.0	74710.0	77050.0	79417.0
1	Africa Eastern and Southern	AFE	309890664.0	318544083.0	326933522.0	335625136.0	344418362.0	353466601.0	362985802.0
2	Afghanistan	AFG	10694796.0	10745167.0	12057433.0	14003760.0	15455555.0	16418912.0	17106595.0

To analyze which countries have the lowest amount of increasing CO2 emissions compared to their population growth, I will perform the following steps using the datasets which are output of my data pipeline.

Define Population Growth: Calculate the population growth for each country. (from 1990 to 2020)

I defined population growth as: $(Population_{2020} - Population_{1990}) / Population_{1990} \times 100$

Define CO2 Emission Increase: Calculate the CO2 emission growth for each country. (from 1990 to 2020) I defined CO2 emission increase as: $(CO2\ Emissions_{2020} - CO2\ Emissions_{1990}) / CO2\ Emissions_{1990} \times 100$

For these definitions, I simply selected population and CO2 emission values of the years 1990 and 2020 to calculate the change over these years.

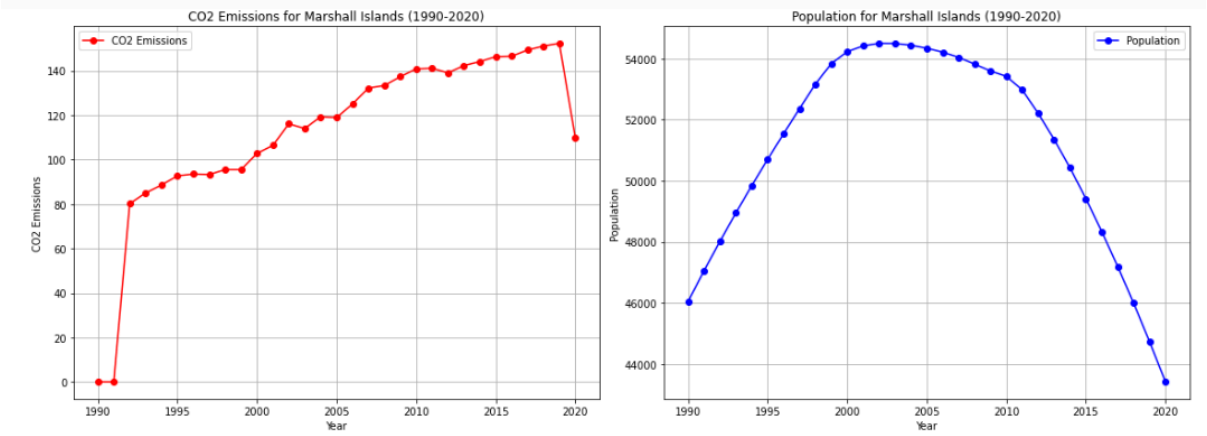
Compare CO2 Emission Growth to Population Growth: Analyse and rank the countries based on the ratio of CO2 emission growth to population growth.

This will identify the countries with the lowest CO2 emissions increase relative to their population growth.

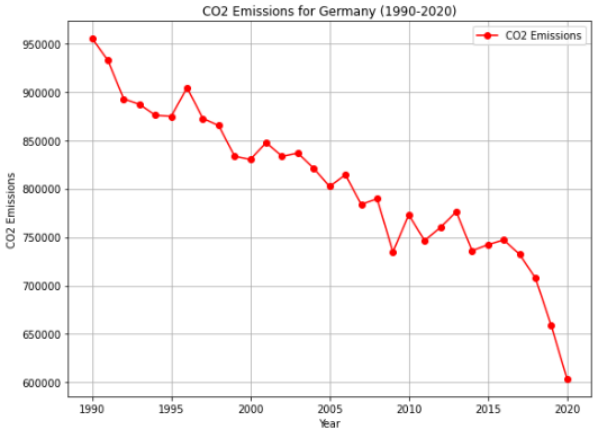
Top 10 countries with the lowest ratio of CO2 emissions growth to population growth:

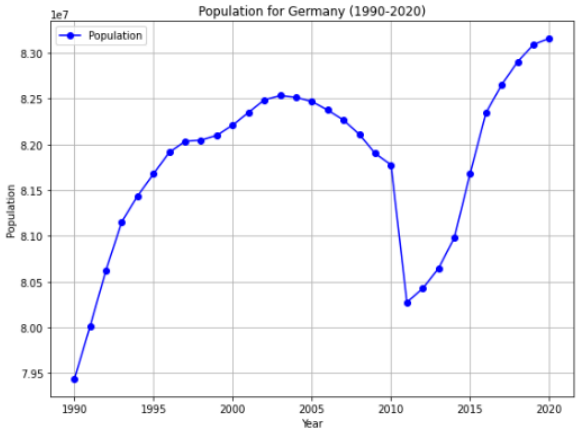
	Country Name	Population_Growth	CO2_Growth	CO2_Population_Growth_Ratio
0	Marshall Islands	-5.72	10900.00	-1905.51
1	St. Vincent and the Grenadines	-6.98	166.10	-23.79
2	Slovak Republic	3.01	-47.03	-15.61
3	Czechia	3.53	-41.99	-11.91
4	Germany	4.69	-36.84	-7.85
5	Greece	4.92	-32.55	-6.62
6	Italy	4.79	-30.59	-6.38
7	Cuba	6.34	-31.70	-5.00
8	European Union	6.52	-30.91	-4.74
9	Denmark	13.43	-47.37	-3.53

In some countries such as the Marshall Islands, the CO2 emission value was recorded as "0" in 1990. This led to errors due to division by zero. To resolve this issue, I opted to replace the zero values with one, which appeared to be the most suitable solution for these specific countries. I believe this solution mitigates any significant data loss mathematically. However, in countries where CO2 emissions were originally very low, even a slight increase results in a large proportional change. There might be a more effective way to represent these incremental increases.



However, when considering other countries on the list such as Germany, European Union and Denmark, they are economically developed with strict environmental regulations, policies and a strong environmental consciousness. Therefore, this list substantially reflects the actual situation in the world.





4. Conclusions

Based on the analysis:

- **Environmental Policies:** Countries with stringent environmental regulations and policies tend to have better control over emissions growth despite population increases.
- **Public Awareness and Behavior Change:** Increased awareness of environmental issues and sustainable practices among the population can lead to reduced energy consumption and emissions.
- **Economic Structure:** Countries with a diverse economy or sectors that are less carbon intensive may naturally have lower emissions growth relative to population growth.

In summary, countries in this list likely benefit from a combination of proactive environmental policies, technological innovation, public engagement in sustainability and favorable economic structures that collectively contribute to managing CO2 emissions growth despite demographic changes. However, to give a direct answer to our main question, we need better ways to calculate population growth and CO2 emission increase.