Final report: CO_2 emissions

This report aims to analyse global CO_2 emissions and related factors using data from "Our World in Data."
The analysis focuses on:
1. Identifying the biggest predictor of CO_2 output per capita.

2. Determining which countries are making significant strides in reducing CO_2 emissions.

3. Evaluating the future cost-effectiveness of non-fossil fuel energy technologies.

By leveraging Python for data cleaning, pre-processing, and analysis, this report provides insights into the key drivers of CO_2 emissions and the progress of various countries in mitigating their environmental impact. The findings will inform policy recommendations and highlight promising energy technologies for a sustainable future.

1. **What is the biggest predictor of a large CO_2 output per capita of a country?**

To answer the first research question, we need to break it down into more specific sub questions:

- What are the key factors potentially influencing CO_2 output per capita?
- How do these factors correlate with CO_2 output per capita?
- Which factor has the strongest correlation with CO_2 output per capita?

The next step is to review existing literature to understand what previous studies have identified as predictors for CO_2 emissions.

- What are the key factors potentially influencing CO_2 output per capita?

According to previous studies and research there are 4 fundamental factors*:

1. Human population
2. Gross Domestic Product (GDP) per capita
3. Energy intensity (per unit of GDP)
4. Carbon intensity (emissions per unit of energy consumed)

*source: https://en.wikipedia.org/wiki/Kaya_identity

- How do these factors correlate with CO_2 output per capita?

Create a correlation matrix heatmap and calculate the Pearson correlation coefficient between the strongest correlation according to the heatmap.

Methodology

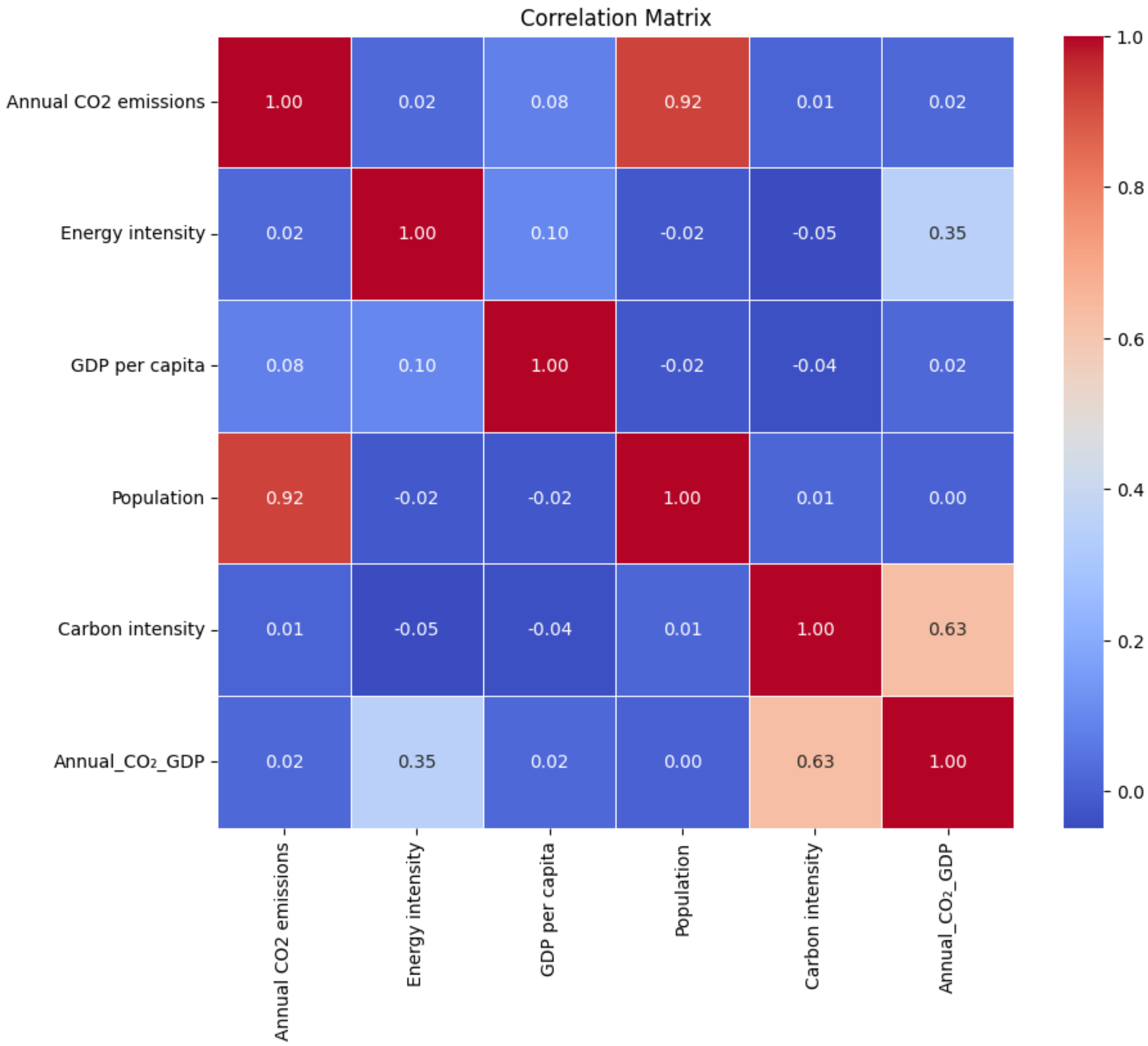
The following steps were taken to achieve the results:

- Data Preparation and Cleaning
  Renamed columns for clarity
- Dropped irrelevant columns:
  Removed the "900793-annotations" column
- Handled missing data:
  Identified rows with missing 'Code' (which were aggregate data for Africa)
  Removed all rows with any missing data using `df.dropna(inplace=True)`
- Exploratory Data Analysis
  1. Displayed the first few rows of the cleaned dataset
  2. Generated descriptive statistics using `df.describe()`
  3. Verified the absence of missing values
- Key Observations:
  The cleaned dataset contains 11,903 rows
  Years range from 1965 to 2021
  Annual CO_2 emissions range from 0 to about 37 billion tons per year
  GDP per capita ranges from about $378 to $160,051

Population ranges from 1,833 to about 7.9 billion

- 3. Correlation Analysis

1. Created a correlation matrix heatmap for all numerical variables
2. Calculated the Pearson correlation coefficient specifically between Population and Annual CO2 emissions



Results:

1. The heatmap revealed various correlations between the variables
2. Pearson correlation coefficient between Population and Annual CO2 emissions: 0.9696
3. P-value for this correlation: 0.0 (extremely small)

The Matrix shows there is a very significant correlation between population and the Annual CO_2 emissions.
This is confirmed by pearsons correlation coeffcient of 0,92 and a P-value of 0.
A positive correlation indicates that two variables will move in the same direction. In other words, if the population increases, the annual CO2 emissions will increase as well, and if one variable decreases the other decreases equivalently.

With a P-value below 0,05 the correlation coefficient is called statistically significant.

Implications of the Results

1. Strong Positive Correlation: The very high Pearson correlation coefficient (0.9696) between Population and Annual CO_2 emissions indicates a strong positive relationship. This suggests that as population increases, CO_2 emissions tend to increase as well, and vice versa.

2. Statistical Significance: The extremely low p-value (0.0) indicates that this correlation is statistically significant. The probability of observing such a strong correlation by chance is essentially zero.

3. Complex Relationships: While the correlation is strong, it's crucial to note that correlation does not imply causation. The relationship between population and CO_2 emissions is likely influenced by various factors not captured in this simple analysis, such as:

- Level of industrialization
- Economic development
- Energy policies and practices
- Technological advancements
- Consumption patterns

4. Policy Implications: The strong relationship between population and CO_2 emissions suggests that population growth could be a significant factor in climate change discussions. However, policymakers should consider this in conjunction with other factors like per capita emissions, technological solutions, and sustainable development practices.

5. Need for Further Analysis: While this correlation provides valuable insights, it also highlights the need for more sophisticated analyses. Multivariate analyses, time series analysis, or regression models could provide a more comprehensive understanding of the factors influencing CO_2 emissions.

In conclusion, this analysis provides a strong foundation for understanding the relationship between population and CO_2 emissions. However, it also opens up avenues for more detailed investigations into the complex factors influencing global CO_2 emissions.

2. **Which countries are making the biggest strides in decreasing CO_2 output?**

The objective of this analysis was to identify the top 10 countries that have achieved the most significant reduction in CO_2 emissions relative to their population change from the year 2000 to 2022. This analysis helps to understand which countries have been most effective in reducing their carbon footprint while accounting for population growth or decline.

The data used for this analysis was obtained from the "Our World in Data" CO_2 dataset, which includes information on CO_2 emissions, population, and other relevant metrics for various countries over multiple years.

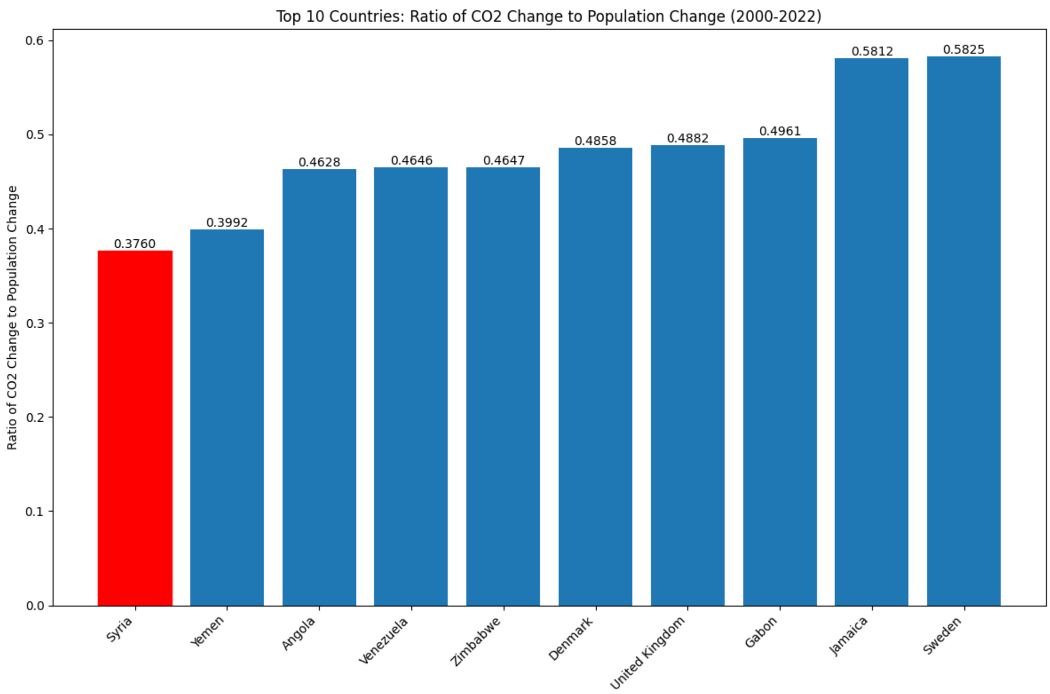
Methodology

The following steps were taken to achieve the results:

- Data for the years 2000 and 2022 was selected to compare the changes over this period.
- Key columns such as 'country', 'year', 'population', and 'co2' were retained for further analysis.
- The data was pivoted to have years as columns, allowing for easy calculation of changes between 2000 and 2022.
- The relative change in population from 2000 to 2022 was calculated using the formula:
  population_change = (population2022 / population2000) × 100
- Similarly, the relative change in CO_2 emissions from 2000 to 2022 was calculated using the formula: co2_change = (co22022 / co22000) × 100
- Countries with extreme population growth or shrinkage were filtered out by setting thresholds for population change (between 70% and 300%).
- Additionally, countries with a population of less than 1,000,000 in the year 2000 were excluded to focus on more significant cases.
- Calculation of CO_2 to Population Change Ratio
  The ratio of CO_2 change to population change was calculated for each country using the formula:
  Ratio = co2_change / population_change
- Selection of Top 10 Countries
  The countries were sorted by the calculated ratio, and the top 10 countries with the lowest ratios (indicating the biggest decrease in CO_2 emissions relative to population change) were selected.
- Visualization
  A bar chart was created to visualize the relative changes in CO_2 emissions and population for the top 10 countries. The chart highlights the ratio of CO_2 change to population change, with the countries sorted from lowest to highest ratio.

Results

The top 10 countries with the biggest CO_2 decrease relative to population change from 2000 to 2022 are as follows:



Interpretation of Results:

Syria has the lowest ratio (0.3760), indicating the most significant decrease in CO_2 emissions relative to its population change. Despite the challenges faced by the country, it has managed to reduce its carbon footprint effectively.

Yemen and Angola follow closely behind Syria, with ratios of 0.3992 and 0.4628, respectively.

Venezuela and Zimbabwe also show significant reductions in CO_2 emissions relative to their population changes, with ratios around 0.4646.

Denmark and the United Kingdom are notable developed countries that have achieved substantial reductions in CO_2 emissions relative to their population growth, with ratios of 0.4858 and 0.4882, respectively.

Gabon, Jamaica and Sweden complete the top 10 list, with ratios ranging from 0.4961 to 0.5825.

Conclusion:

This analysis highlights the countries that have been most effective in reducing their CO_2 emissions relative to their population changes from 2000 to 2022. The results provide valuable insights into the efforts and policies implemented by these countries to combat climate change. Further research could explore the specific measures taken by these countries to achieve such reductions and how they can be applied to other regions.

3. **Which non-fossil fuel energy technology will have the best price in the future?**

The data used for this analysis was obtained from the "Our World in Data" Energy datasets, which includes information on different types of renewable energy and the according prices in USD per KWh.

Methodology
The following steps were taken to achieve the results:

- Data Loading
  Two CSV files were loaded: 'energy.csv' and 'levelized-cost-of-energy.csv' from a GitHub repository.
  The data was loaded into pandas DataFrames named df3 and df4 respectively.
- Data Filtering and Preprocessing
  A new DataFrame 'world_energy' was created by filtering df4 for rows where 'Entity' is 'World'.
  Columns were renamed for clarity and irrelevant columns ('code', 'csp', 'country') were dropped.
  The DataFrame was sorted by year.
- Initial Data Inspection
  The head of the DataFrame was examined.
  Basic information about the DataFrame structure was obtained.
  Summary statistics were generated.
  Missing values were checked.
  The range of years in the dataset was identified (1983 to 2019).
- Data Visualization
  Initial Plotting
  A single plot was created showing the price trends of all renewable energy types from 2000 to 2020.
  This plot provided an overview of how different energy prices compared and evolved over time.
- Detailed Individual Plots
  A function 'plot_regression' was created to generate individual plots for each energy type.
  This function:
  a. Filters the data for the specific energy type.
  b. Plots the actual price data points.
  c. Calculates and plots a linear regression line to show the trend.
  d. Sets appropriate labels and titles.
  The function was applied to each energy type: hydro, solar, onshore wind, biomass, geothermal, and offshore wind.
- Results and Interpretation
  The analysis reveals several key insights about renewable energy price trends:
  1. Overall Trend: All renewable energy sources show a downward price trend over time, indicating improving cost-effectiveness and technological advancements across the board.

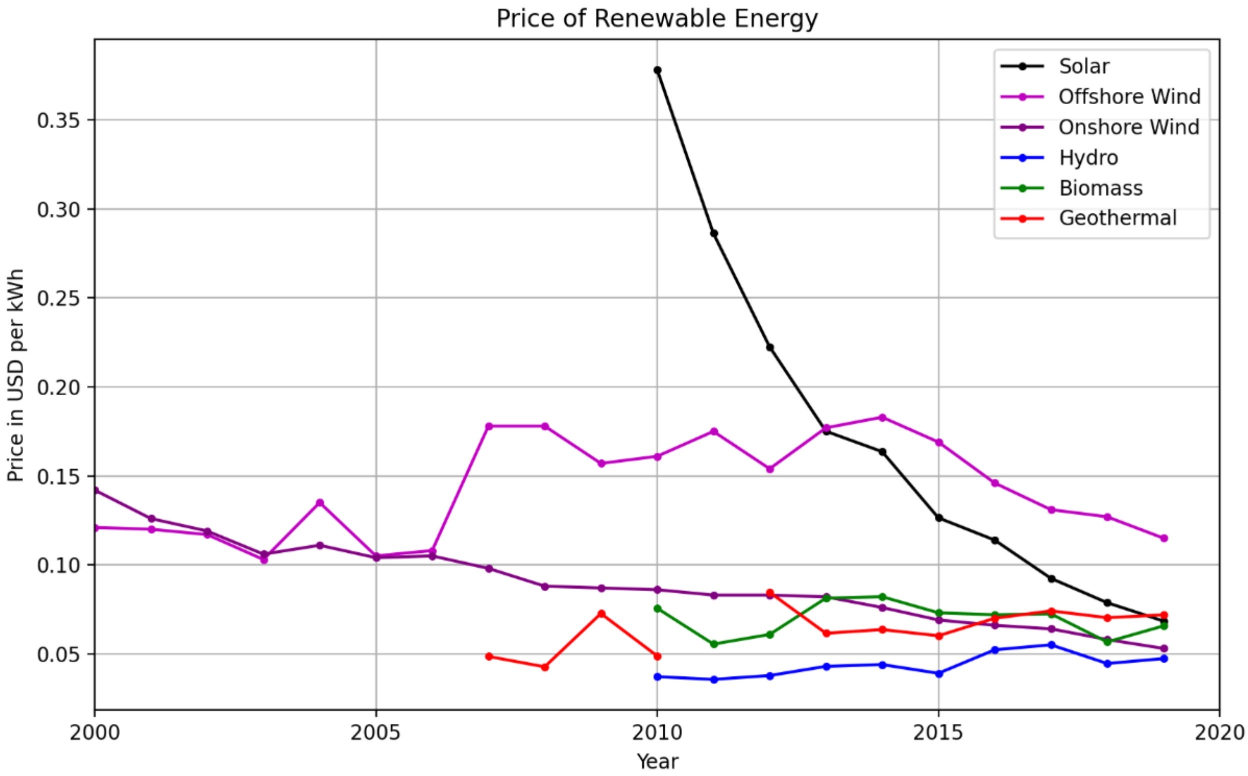
  2. Solar Energy: Demonstrates the most dramatic price decrease. From being one of the most expensive options in the early 2000s, it has become one of the cheapest by 2020. This rapid decline suggests significant technological improvements and economies of scale in solar panel production.

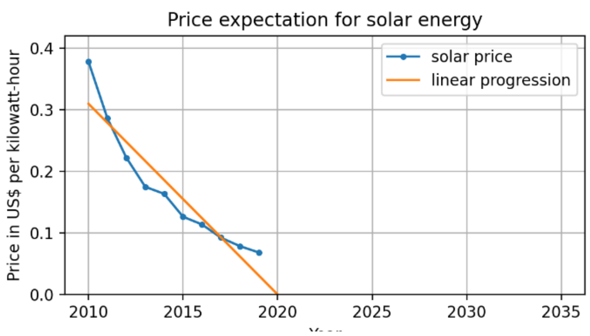
  3. Wind Energy:
  - Onshore Wind: Shows a significant price reduction, though not as steep as solar. It has been consistently decreasing in price since the 1980s.
  - Offshore Wind: Despite data starting later, it shows a clear and significant downward trend, indicating rapid advancements in offshore wind technology.

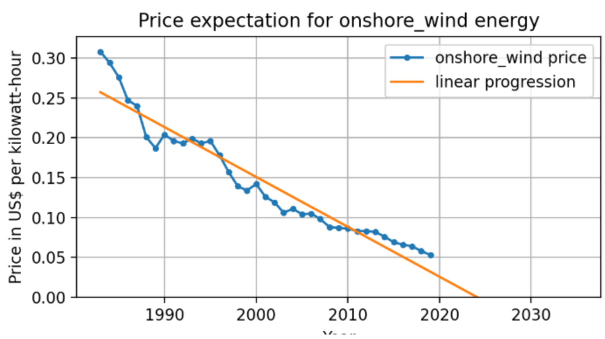
  4. Hydro Energy: Exhibits a slight downward trend with low volatility. This suggests a mature technology with gradual improvements over time.

  5. Biomass Energy: Shows a modest downward trend. The clustered data points indicate relatively stable prices with gradual improvements in cost-effectiveness.

  6. Geothermal Energy: Demonstrates a gradual decrease in prices, suggesting steady but slower technological advancements compared to solar and wind.







Implications:
1. Increasing Competitiveness: The declining costs across all renewable sources, especially solar and wind, indicate that these technologies are becoming increasingly competitive with traditional fossil fuel energy sources.

2. Technological Progress: The steep decline in solar and wind energy prices reflects rapid technological advancements in these sectors.

3. Investment Trends: These trends may influence future investment decisions, potentially favoring solar and wind energy projects due to their rapidly decreasing costs.

4. Policy Implications: The increasing cost-effectiveness of renewables could support more aggressive clean energy policies and faster transitions away from fossil fuels.

5. Future Energy Mix: As costs continue to decrease, we might expect to see a greater share of solar and wind in the global energy mix, complemented by other renewable sources.

6. Research and Development: While solar and wind show the most dramatic improvements, the consistent downward trends across all types suggest ongoing R&D efforts are yielding results across the renewable energy sector.

This analysis provides a clear picture of the evolving economics of renewable energy, highlighting the increasing viability of these technologies as primary energy sources for the future. Based on the results, Solar-and onshore wind energy will have the best price in the future.