Final Assignment Report

Analysis of CO_2 Emissions and Related Factors

Nicolette Verleur

Introduction

This report represents the culmination of our journey through the "Data Analytics with Python" course. As we delve into the world of data analysis, we turn our attention to three fundamental questions related to CO_2 emissions. Our exploration serves as a practical application of the coding skills we've acquired.

The questions we address are:

1. What is the biggest predictor of a large CO_2 output per capita of a country?
2. Which countries are making the biggest strides in decreasing CO_2 output?
3. Which non-fossil fuel energy technology will have the best price in the future?

These questions touch on significant environmental and economic concerns. By utilizing data analysis techniques, we seek to extract insights that contribute to our understanding of these issues.

Throughout this report, we will delve into these questions and present the insights derived from our analysis.

Methodology

Data Sources

The primary data sources for our analysis are derived from the 'Our World in Data' website. We obtained relevant datasets from this source to address the questions presented in this report. These datasets encompass data related to CO_2 emissions, non-fossil fuel energy technologies, and potential predictor variables.

Question 1: Biggest predictor of CO_2 output per capita

Approach

For this question, we investigated potential predictors of high CO_2 emissions per capita using correlation analysis (Pearson). The dataset, 'co-emissions-per-capita.csv', provided detailed information about CO_2 emissions per capita. A range of potential predictors, were derived from multiple CSV files, containing one or more potential predictors. The list of potential predictors can be seen in Table 1. The CSV file 'new-vehicles-type-area.csv' was excluded from the analysis, due to its limited sample size (309 rows). The CSV files were merged together (after preprocessing), with an inner merge on the columns 'Country' and 'Year', resulting in our final dataset involving 2300 rows, 170 countries, and 18 years (from 2000 to 2017). This dataset contains a few minor instances of missing values in certain columns. However, as these missing values account for less than 5.4%, we chose not to remove the corresponding rows.

column	Interpretation
Country	Geographic location
Year	Year of observation
cement_co2_per_capita	Annual production-based emissions of carbon dioxide (CO_2) from cement, measured in tonnes per person.
	This is based on territorial emissions, which do not account for emissions embedded in traded goods.
co2_per_capita	Annual total production-based emissions of carbon dioxide (CO_2), excluding land-use change, measured in tonnes per person.
	This is based on territorial emissions, which do not account for emissions embedded in traded goods.",Global Carbon Budget (2022)
coal_co2_per_capita	Annual production-based emissions of carbon dioxide (CO_2) from coal, measured in tonnes per person.
	This is based on territorial emissions, which do not account for emissions embedded in traded goods.
energy_per_capita	Primary energy consumption per capita, measured in kilowatt-hours per person per year.
energy_per_gdp	Primary energy consumption per unit of gross domestic product, measured in kilowatt-hours per international-dollar.
flaring_co2_per_capita	Annual production-based emissions of carbon dioxide (CO_2) from flaring, measured in tonnes per person.
	This is based on territorial emissions, which do not account for emissions embedded in traded goods.
Food supply (kcal per capita per day)	Daily per capita caloric supply is measured in kilocalories per person per day.
	This indicates the caloric availability delivered to households but does not necessarily indicate the number of calories actually consumed.
	This measures the quantity that is available for consumption at the end of the supply chain.
	It does not account for consumer waste, so the quantity that is actually consumed may be lower than this value.
Food supply (Protein g per capita per day)	Average daily per capita protein supply, measured in grams of total protein per day.
	This measures the quantity that is available for consumption at the end of the supply chain.
	It does not account for consumer waste, so the quantity that is actually consumed may be lower than this value.
Food supply (Fat g per capita per day)	Average daily per capita dietary fat supply, measured in grams per person per day.
	This measures the quantity that is available for consumption at the end of the supply chain.
	It does not account for consumer waste, so the quantity that is actually consumed may be lower than this value.
Food (CV of caloric intake)	The inquality in dietary calorie intake is measured as the coefficient of variation (CV) of energy intake.
	It represents the spread of intakes around the mean.
	Higher CV values represent larger levels of dietary inequality.
fossile_co2_per_capita	Sum of the columns coal_co2_per_capita, 'gas_co2_per_capita and oil_co2_per_capita.
	(all in Annual production-based emissions of carbon dioxide (CO_2) from coal, gas and oil, measured in tonnes per person.
	This is based on territorial emissions, which do not account for emissions embedded in traded goods.)
gas_co2_per_capita	Annual production-based emissions of carbon dioxide (CO_2) from gas, measured in million tonnes.
	This is based on territorial emissions, which do not account for emissions embedded in traded goods.
land_use_change_co2_per_capita	Annual production-based emissions of carbon dioxide (CO_2) from land-use change, measured in tonnes per person.
Meat supply (kg per year per capita)	Meat supply (kg per year per capita)
	This measures the quantity that is available for consumption at the end of the supply chain.
methane_per_capita	Total methane emissions including land-use change and forestry, measured in tonnes of carbon dioxide-equivalents per capita.
new_vehicles_battery_electric_(number/capita)	Annual new vehicles battery electric_(number per capita)
new_vehicles_plugin_hybrid_(number/capita)	Annual new vehicles plugin hybrid (number per capita)
new_vehicles_full_mild_hybrid_(number/capita)	Annual new vehicles full mild hybrid (number per capita)
new_vehicles_petrol_(number/capita)	Annual new vehicles petrol (number per capita)
new_vehicles_diesel_gas_(number/capita)	Annual new vehicles diesel gas(number per capita)
new_vehicles_(number/capita)	Sum of annual new vehicles on: battery electric, plugin hybrid, full mild hybrid, petrol and diesel gas(number per capita)
nitrous_oxide_per_capita	Total nitrous oxide emissions including land-use change and forestry, measured in tonnes of carbon dioxide-equivalents per capita.
oil_co2_per_capita	Annual production-based emissions of carbon dioxide (CO_2) from oil, measured in tonnes per person.
	This is based on territorial emissions, which do not account for emissions embedded in traded goods.
Population density (people/km2)	Population density (people/km2)
Human Development Index (HDI)	The Human Development Index (HDI) is a summary measure of key dimensions of human development:
	a long and healthy life, a good education, and having a decent standard of living.
	(a measure of human development that captures health, education, and income)
Average Total Years of Schooling	Average Total Years of Schooling for Adult Population (Lee-Lee (2016), Barro-Lee (2018) and UNDP (2018))
Life_expectancy_at_birth	Life expectancy at birth (historical)
Expected Years of Schooling	Expected Years of Schooling
Average Total Years of Schooling	Average Total Years of Schooling for Adult Population (Lee-Lee (2016), Barro-Lee (2018) and UNDP (2018))
GDP_per_capita	Gross domestic product (GDP) measured in international-dollar using 2011 prices to adjust for price changes over time (inflation) and price differences between countries.
GNI_per_capita	Gross national income (GNI) measures the total income earned by residents of a country, including income earned abroad.
	This data is adjusted for inflation and differences in the cost of living between countries.

Tabel 1: list of columns, representing potential predictors of high CO_2 emissions per capita. ('country' and 'year' are the columns on which the merge was conducted, and 'co2_per_capita' is the column against which the other predictors' correlation coefficients were compared in the data analysis. The columns starting with 'new_vehicles' where not included in the final data frame, due to limited sample size.

Findings

Among the predictors studied, we found a strong positive correlation between CO_2 emissions per capita and the consumption of fossil fuels (gas, oil, and coal) with a remarkable correlation coefficient of 0.999. This coefficient surpassed the individual correlations of the three separate energy sources, as shown in Table 2. This phenomenon could be attributed to the varying combinations of energy forms that different countries adopt.

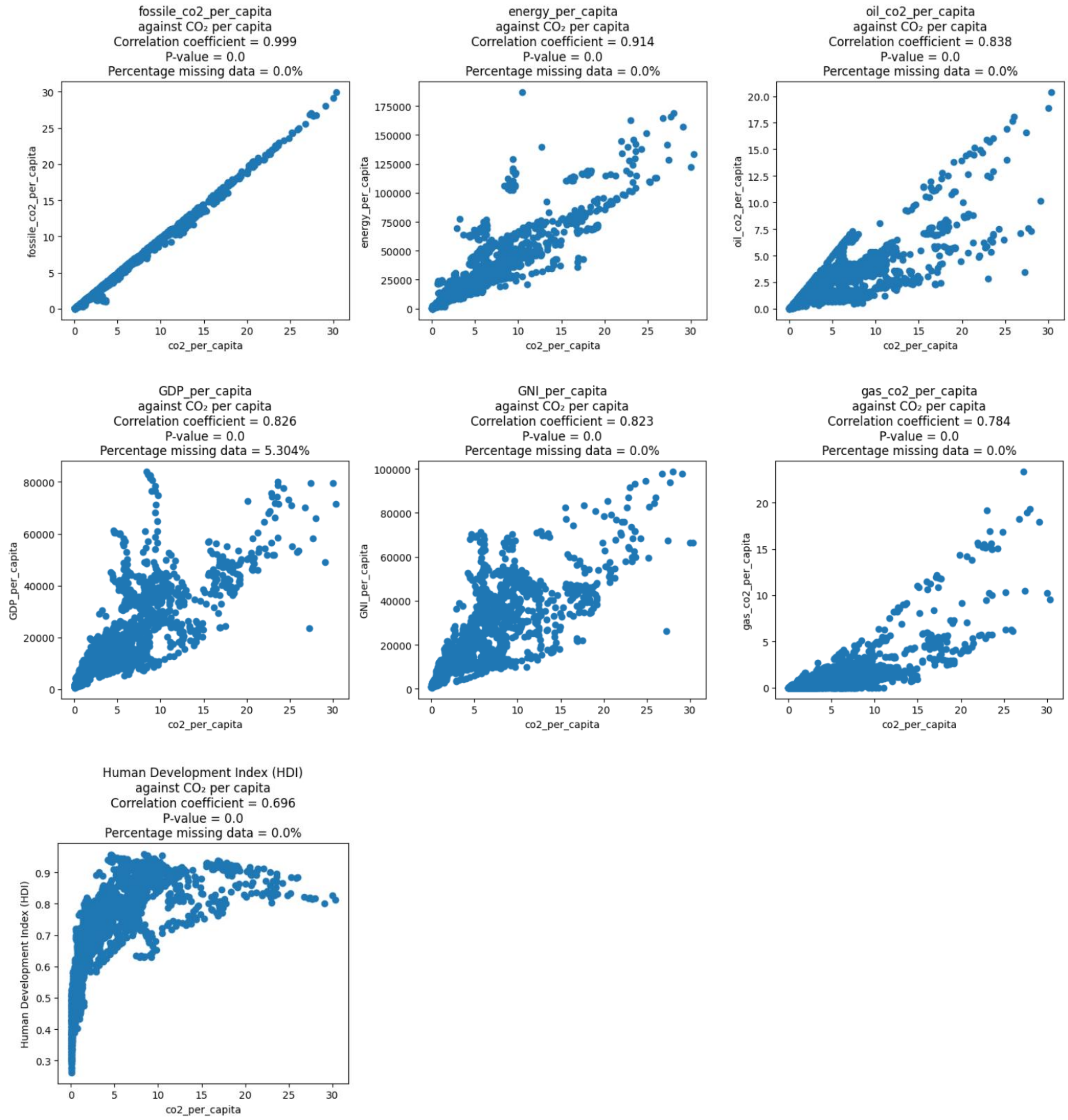
Furthermore, primary energy consumption per capita exhibited a substantial correlation of 0.9138, indicating its role as a predictor. Indicators of economic strength, such as Gross Domestic Product (GDP) per capita and Gross National Income (GNI) per capita, demonstrated strong predictive capabilities with correlation coefficients of 0.8259 and 0.8233, respectively.

The Human Development Index (HDI), which reflects key dimensions of human development, held a correlation coefficient of 0.6958, indicating its relevance as a predictor.

	corr_c_co2_per_capita	p_value_co2_per_capita	Percentage_missing
co2_per_capita	1.000000	0.0	0.000000
fossile_co2_per_capita	0.998841	0.0	0.000000
energy_per_capita	0.913809	0.0	0.000000
oil_co2_per_capita	0.838077	0.0	0.000000
GDP_per_capita	0.825908	0.0	5.304348
GNI_per_capita	0.823373	0.0	0.000000
gas_co2_per_capita	0.783589	0.0	0.000000
Human Development Index (HDI)	0.695788	0.0	0.000000
Meat supply (kg per year per capita)	0.660612	0.0	0.000000
Food supply (Fat g per capita per day)	0.647807	0.0	0.000000
cement_co2_per_capita	0.636585	0.0	1.260870
Food supply (kcal per capita per day)	0.629102	0.0	0.000000
Food supply (Protein g per capita per day)	0.627157	0.0	0.000000
Average Total Years of Schooling	0.624433	0.0	0.000000
Expected Years of Schooling	0.602058	0.0	0.000000
coal_co2_per_capita	0.585744	0.0	0.000000
Life_expectancy_at_birth	0.576766	0.0	0.000000
energy_per_gdp	0.492474	0.0	5.304348
methane_per_capita	0.416131	0.0	0.782609
flaring_co2_per_capita	0.330653	0.0	0.000000
nitrous_oxide_per_capita	0.146276	0.0	0.782609
Population density (people/km2)	0.022304	0.285	0.000000
land_use_change_co2_per_capita	-0.193269	0.0	1.043478
Food (CV of caloric intake)	-0.538379	0.0	0.000000

Tabel 2: Pearson correlation coefficients, p-values and missing values of several possible predictors in comparison with the CO_2 emission per capita.

The scatter plots from the predictors with correlation coefficients >= 0.69 are shown in the plots below. It is noteworthy that the reliability of Pearson correlation coefficients is highest when variables follow a normal distribution and exhibit a linear relationship. However, in this analysis, the normality of the distribution was not investigated. In the case of the 'fossil energy' predictor, a very strong linear relationship is evident. As the linear relationship weakens for other predictors, the correlation coefficient also decreases.



Scatter plots from the predictors with correlation coefficients >= 0.69

Question 2: Biggest strides in decreasing CO_2 output

Approach and Key Findings Summary

In our pursuit to identify countries making significant strides in reducing CO_2 emissions, we explored three distinct approaches. Each approach offered different top-performing countries based on varying criteria. During data preparation, rows with missing values in the 'Code' column were excluded, along with any rows containing the entry 'World' under the 'Entity' column if present.

Analysis of Top 10 countries with the highest percentage reduction in CO_2 emissions per capita

Approach Q2a - Decrease in relative CO_2 emission in 2020 compared to 1990

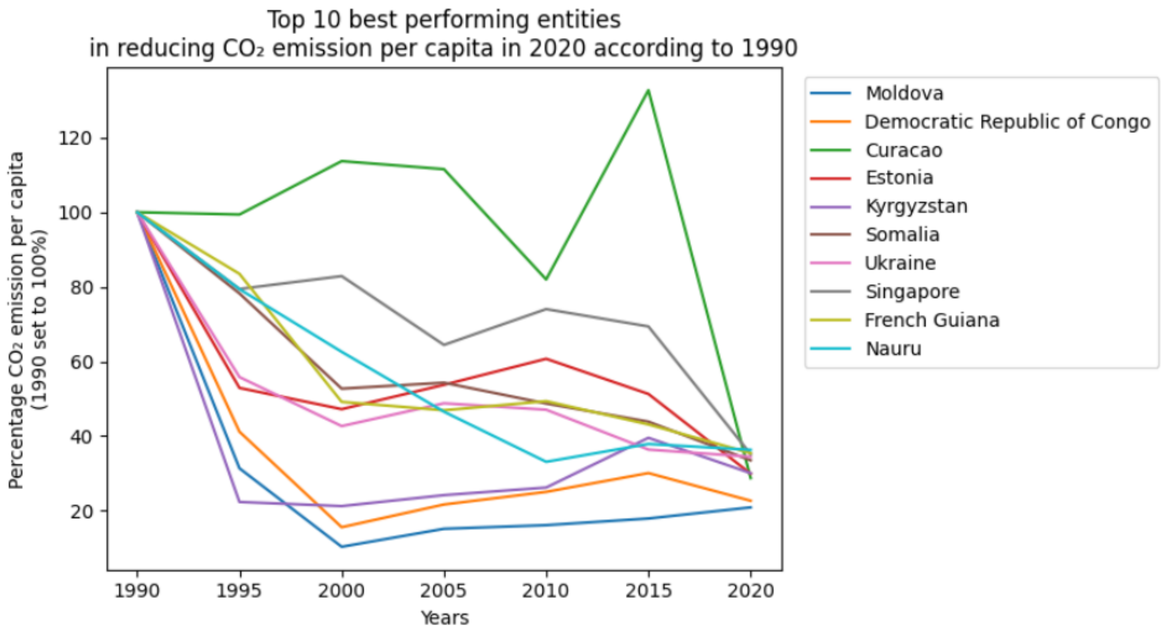
Our initial exploration focused on evaluating the most substantial percentage decrease in CO_2 emissions per capita between 2020 and 1990. This perspective aligned with the commitments set forth in the 2015 Paris Climate Agreement, guiding efforts towards a 55% reduction by 2030.

The data was selected for specific years (1990-2020, interval 5 years) and transformed (pivot) into a structured dataframe. The emissions of 1990 served as the baseline with a value of 100%. The top 10 countries with the most significant percentage reductions of CO_2 emissions per capita were identified and presented in a graph.

The top 10 countries with the greatest reductions were as follows:

1. Moldova
2. Democratic Republic of Congo
3. Curacao
4. Estonia
5. Kyrgyzstan
6. Somalia
7. Ukraine
8. Singapore
9. French Guiana
10. Nauru

Interestingly, these countries exhibited varying patterns of progress, with some showing consistent decreases over the years, while others experienced fluctuations in their reduction trajectories. For instance Curaçao is earning its spot in this top 10 for its efforts in recent years, while other countries totally rely on their efforts in the past (Moldavia, Democratic Republic of Congo and Kyrgyzstan). Somalia and Singapore showed more consistent progress.

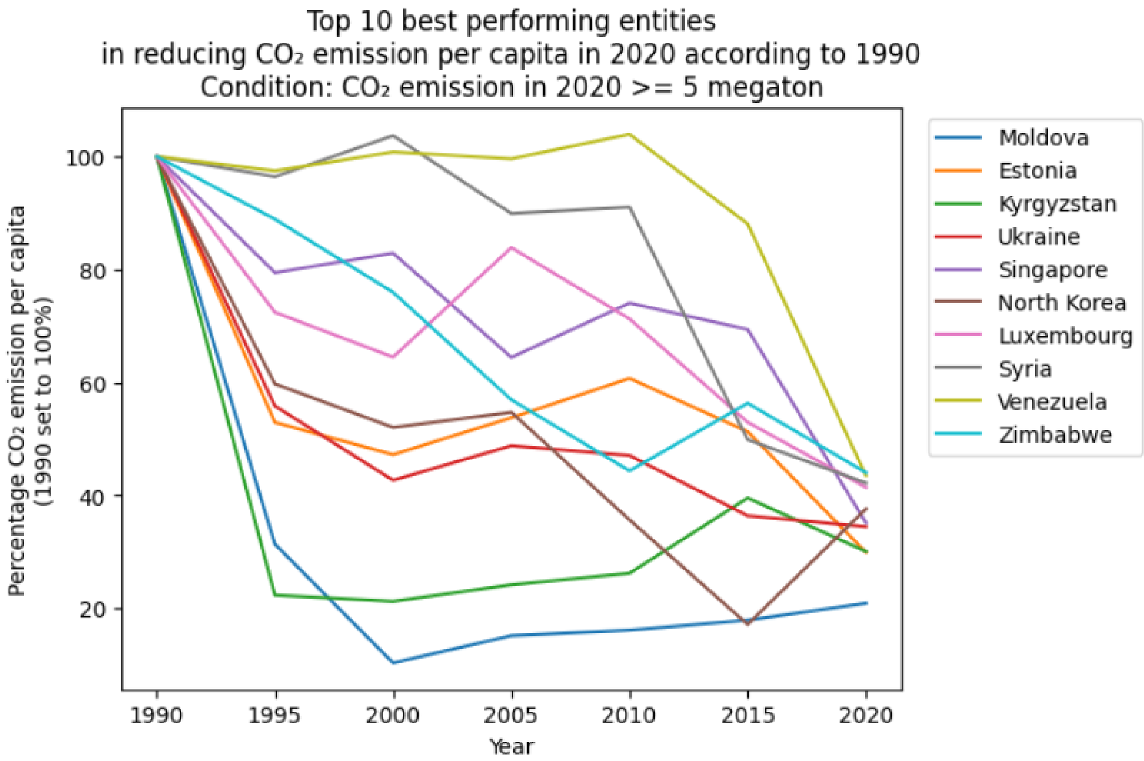


Plot 1: Top 10 best performing entities in reducing CO_2 emission per capita in 2020 according to 1990. In the legend, the entities are listed in order of the top 10, with the country having the largest decrease first.

Approach Q2b - Decrease in relative CO_2 emission in 2020 compared to 1990, with a variable condition for minimal CO_2 emission in 2020

Building on the previous approach, a more specific analysis was conducted by considering countries with minimal CO_2 emissions in 2020. A function was developed to allow the flexibility of setting a threshold for CO_2 emissions in 2020. This new function filtered countries meeting the threshold and produced a graph showcasing the top 10 entities with the largest emissions reductions relative to 1990, with the condition that only countries with a minimal CO_2 emissions in 2020 were considered.

By setting a minimum emission threshold of 5 megatons of CO_2 emissions in 2020 (notably CO_2 emissions, not CO_2 emissions per capita), a different top 10 emerges. Democratic Republic of Congo, Curacao, Somalia, French Guiana and Nauru are no longer part of the top 10. Instead, North Korea, Luxembourg, Syria, Venezuela and Zimbabwe respectively claim positions 6 through 10. These countries also exhibit varying patterns in their achievements over the years.



Plot 2: Top 10 best performing entities in reducing CO_2 emission per capita in 2020 according to 1990. Condition: CO_2 emission in 2020 >= 5 megaton. In the legend, the entities are listed in order of the top 10, with the country having the largest decrease first.

Approach Q2c - Recent Reductions (last 1, 2, 3, 5, and 10 years) *in CO_2 Emissions relative to 2021*

This approach focused on assessing recent reduction percentages in CO_2 emissions per capita over varying time frames (last 1, 2, 3, 5, and 10 years) relative to 2021. The data was selected for specific years and transformed (pivot) into a structured dataframe. The emissions of 2021 served as the baseline with a value of 100%. A custom function was utilized to sort the data and visualize the top 10 performing entities for each time frame, highlighting their reduction trajectories.

For clarity, only the top 3 for each timeframe is given below:

Approach Q2c - Recent Reductions (last year) in CO_2 Emissions relative to 2021:

1. Bosnia and Herzegovina
2. Kosovo
3. Serbia

Approach Q2c - Recent Reductions (last 2 years) in CO_2 Emissions relative to 2021:

1. Bosnia and Herzegovina
2. Kosovo
3. Serbia

Approach Q2c - Recent Reductions (last 3 years) in CO_2 Emissions relative to 2021:

1. Curaçao
2. Estonia
3. Bosnia and Herzegovina

Approach Q2c - Recent Reductions (last 5 years) in CO_2 Emissions relative to 2021:

1. Curaçao
2. Venezuela
3. Estonia

Approach Q2c - Recent Reductions (last 10 years) in CO_2 Emissions relative to 2021:

1. Curaçao
2. Aruba
3. Equatorial Guinea

Discussing the top 3 countries within each time frame provides a focused insight into their achievements:

- Curaçao: Leading the way with significant reductions in CO_2 emissions per capita, Curaçao claimed the number one spot in the last 3, 5, and 10 years. This remarkable decline spanned from 2015 to 2019. Curaçao also managed to keep the CO_2 emissions per capita in 2020 and 2021 on the level of 2019
- Bosnia and Herzegovina, Kosovo, and Serbia: These three countries secured positions in the top 3 for the last year and the last 2 years, while also maintaining a presence in the top 10 for the past 3 and 5 years. However, their inclusion in the top lists is primarily due to the substantial emission reduction observed in 2021, the latest year with available data.
    - It is important to approach these findings with caution, as the dramatic reduction in emissions for these three countries in 2021 may be influenced by factors such as changes in measurement methods or other external influences. A single year's data may not necessarily reflect a consistent trend, and further analysis is needed to determine the sustainability of these reductions over time.
- Aruba: Earning a notable second place in the top 10 for the last 10 years, Aruba's achievement was only driven by a substantial reduction during 2011 to 2013. However, the island's emissions have shown intriguing patterns, with levels remaining relatively constant and on the same level during 1990-1999 and 2013-2021, while witnessing a massive plateau spike 2.5 times the levels of 1990-1999 and 2013-2021 between 2000 and 2011 (see appendix).
- Equatorial Guinea: Demonstrating consistent commitment, Equatorial Guinea displayed a persistent decrease in CO_2 emissions per capita throughout 2011 to 2021. It secured the third position for the 10-year timeframe and the fourth for the 5-year interval.
- Venezuela: With a similar commitment, Venezuela showcased continuous emissions reduction from 2013 to 2020. This earned it second place for the 5-year period and fourth for the 10-year span.
- Estonia: A reduction in CO_2 emissions per capita materialized specifically during 2018-2020 for Estonia, positioning it third for the 5-year period, seventh for the 2-year interval, and ninth for the 10-year timeline.

Complexity and Interpretation

The process of determining a definitive list of top-performing countries is intricate. Different criteria can yield varying results, highlighting the diverse nature of the analysis. It's important to consider several factors when evaluating the performance of countries in reducing emissions.

In some cases, remarkable results are achieved within a single year. However, the accuracy of these outcomes might be influenced by measurement errors or other external factors, making the results, particularly for Bosnia and Herzegovina, Kosovo, and Serbia, somewhat uncertain.

It's noteworthy that some countries might appear in top lists due to achievements from the distant past, such as Moldova and the Democratic Republic of Congo. Conversely, others exhibit significant achievements in a short span, as seen with Curaçao's substantial decline in CO_2 emissions per capita during 2015-2019. Further investigation into the strategies employed by these countries could provide valuable insights.

Additionally, the pace of emission reduction varies. For instance, Somalia and Singapore display a gradual decline, while Curaçao's reduction was rapid. These diverse patterns emphasize that the outcomes of top-performing countries strongly depend on the specific conditions being compared, which greatly influence the results.

In summary, interpreting the results requires a comprehensive understanding of the complexities involved, including the varying criteria, fluctuating patterns, and specific timeframes. This underscores the need to approach the analysis with a holistic perspective, recognizing that different methodologies can lead to different conclusions.

Recommendations for Presentation

To enhance presentation accuracy, we recommend displaying graphs with data points for each year, providing a clearer representation of emission trajectories. This adjustment would allow for more precise observations of countries' progress patterns.

Question 3: Best future price for non-fossil fuel energy

Approach and Methodology

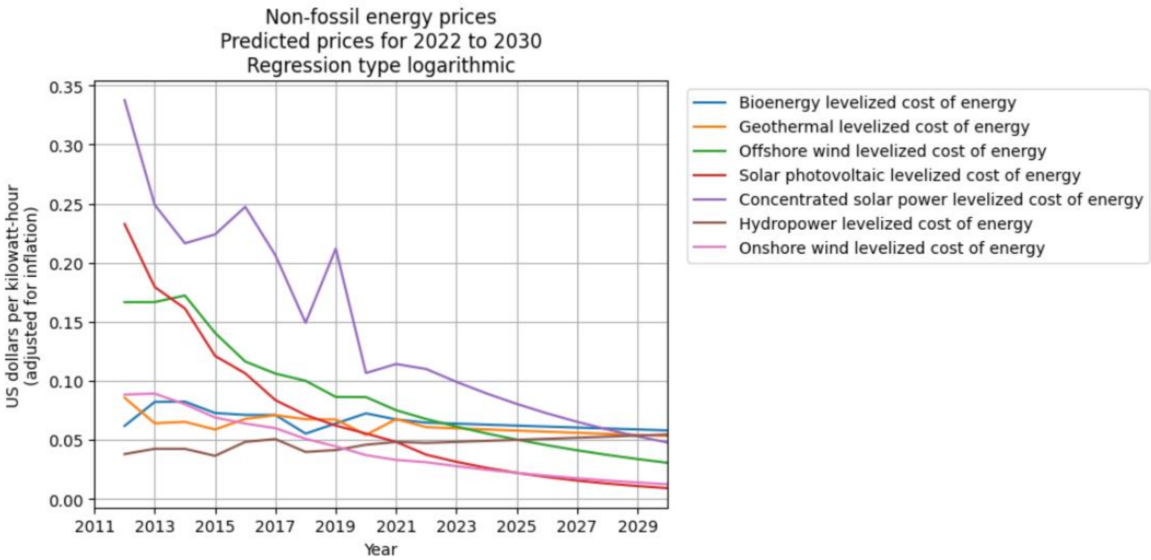
Our exploration into predicting the future prices of various non-fossil fuel energy technologies relied on data derived from the "levelized-cost-of-energy.csv" file from the 'Our World In Data' platform. This dataset encompasses historical prices of various non-fossil fuel energy technologies globally, spanning from 2012 to 2021.

To analyse and forecast these prices for the years 2022 to 2030, we formulated a function integrating multiple regression techniques. This comprehensive approach included linear regression, polynomial regression with degrees 2 and 3, and logarithmic regression. By applying these techniques, we sought to uncover insights into which regression model could accurately predict the most favourable prices for future non-fossil fuel energy technologies.

Findings

The logarithmic regression model provided the most accurate predictions. However, it should be noted that even with this model, there was a slight deviation in the plot between the years 2021 and 2022, as indicated by the red line, which represents the 'Solar photovoltaic levelized cost of energy'.

Among the studied non-fossil energy sources, photovoltaic solar energy and onshore wind energy showed the most promising price trends for the future. These technologies are expected to offer competitive prices compared to other non-fossil fuel energy sources.



Plot 3: Historical non-fossil energy prices(2012 to 2020), and predicted prices for 2022 to 2030, using logarithmic regression.

Considerations and Future Research

It is important to consider that the accuracy of the predictions is subject to the regression model used. Further research and analysis are recommended to validate these findings and assess additional factors that may impact future energy prices.

Conclusions

In the context of understanding the factors contributing to high CO_2 emissions per capita, our analysis has highlighted the pivotal role of fossil fuel consumption. When compared to other influencing factors such as economic and developmental indicators, the consumption of fossil fuels emerges as the strongest predictor. This emphasizes the urgent need for targeted interventions and policy shifts towards more sustainable energy sources to curb emissions effectively.

Turning our attention to countries making significant strides in reducing CO_2 emissions per capita, our exploration encompassed diverse analytical approaches. This approach offered a comprehensive perspective, shedding light on both historical and recent advancements in emission reduction trajectories. The intricacies of our findings underscore the multifaceted nature of emission reduction efforts. Notably, when examining the time span of 2015-2019, Curaçao stands out as an exemplar, demonstrating a remarkable reduction by a factor of 3.5 during these years. This achievement showcases the potential for rapid and substantial emissions decline. It becomes increasingly clear that a holistic assessment, considering various analytical viewpoints, is vital for a thorough understanding of countries' endeavours to combat climate change.

In the realm of predicting the future prices of non-fossil fuel energy technologies, our study harnessed historical data and employed various regression techniques. Through meticulous analysis, logarithmic regression emerged as a robust predictor. Notably, our investigation revealed promising trends for "photovoltaic solar energy" and "onshore wind energy," suggesting competitive and encouraging price trajectories. Nevertheless, the dynamic landscape of energy pricing necessitates ongoing research and continuous exploration to fortify the validity of our conclusions.

In summation, our findings underscore the pivotal role of fossil fuel consumption as a predictor of high CO_2 emissions per capita. The complexity of emission reduction efforts reinforces the value of multi-faceted analytical approaches. Moreover, while our study suggests positive trends for "photovoltaic solar energy" and "onshore wind energy,", the evolving nature of the energy sector demands persistent research to ensure the accuracy and reliability of our predictive insights.

References

Hannah Ritchie, Max Roser and Pablo Rosado (2020) - "CO_2 and Greenhouse Gas Emissions". Published online at OurWorldInData.org. Retrieved from: 'https://ourworldindata.org/co2-and-greenhouse-gas-emissions' [Online Resource]

Max Roser, Hannah Ritchie and Pablo Rosado (2013) - "Food Supply". Published online at OurWorldInData.org. Retrieved from: 'https://ourworldindata.org/food-supply' [Online Resource]

Max Roser (2014) - "Human Development Index (HDI)". Published online at OurWorldInData.org. Retrieved from: 'https://ourworldindata.org/human-development-index' [Online Resource]

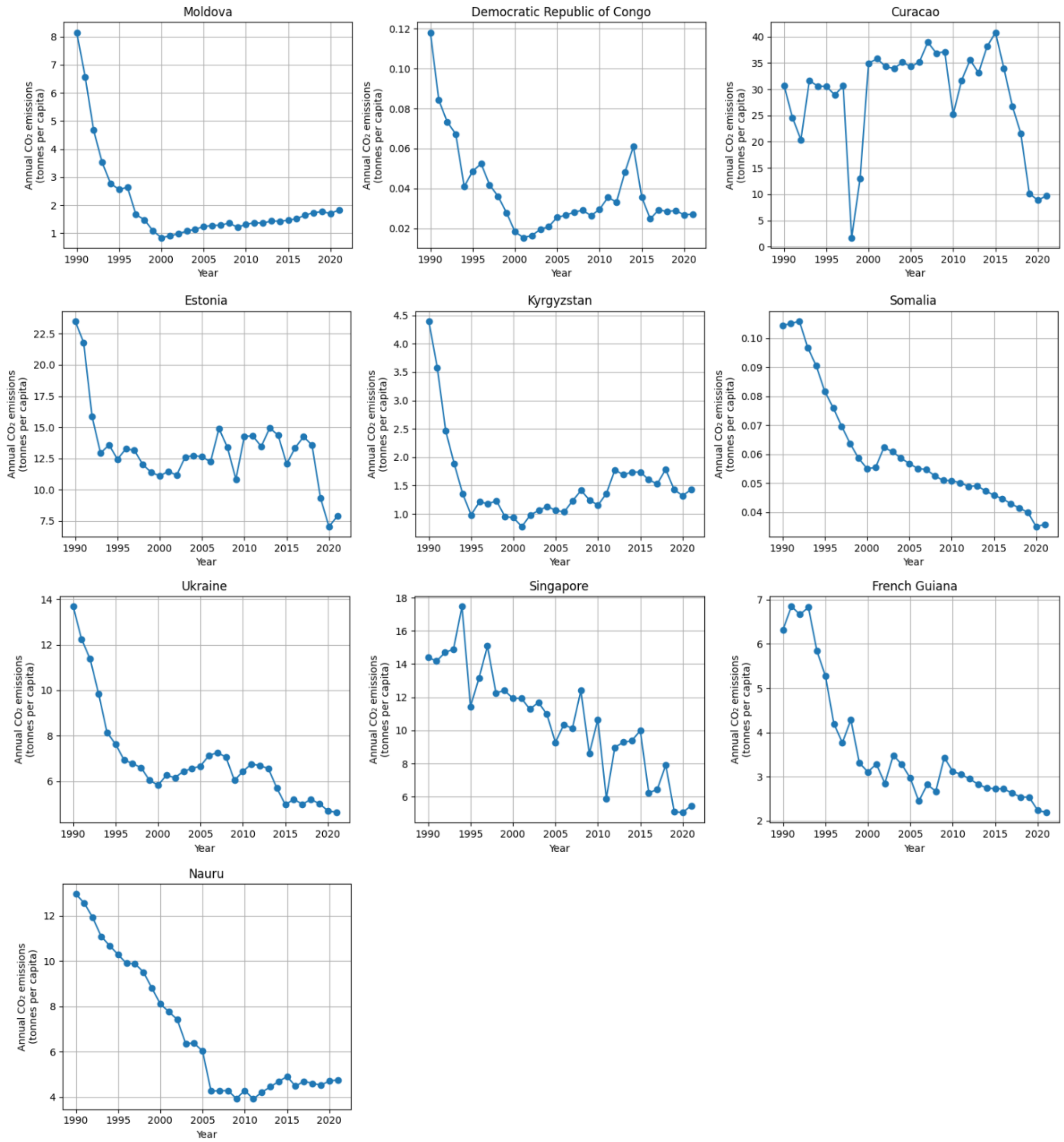
Hannah Ritchie, Max Roser and Pablo Rosado (2022) - "Energy". Published online at OurWorldInData.org. Retrieved from: 'https://ourworldindata.org/energy' [Online Resource]

Interpretation Pierson correlation coefficients: https://towardsdatascience.com/eveything-you-need-to-know-about-interpreting-correlations-2c485841c0b8 (credits to Parvez Ahammad)

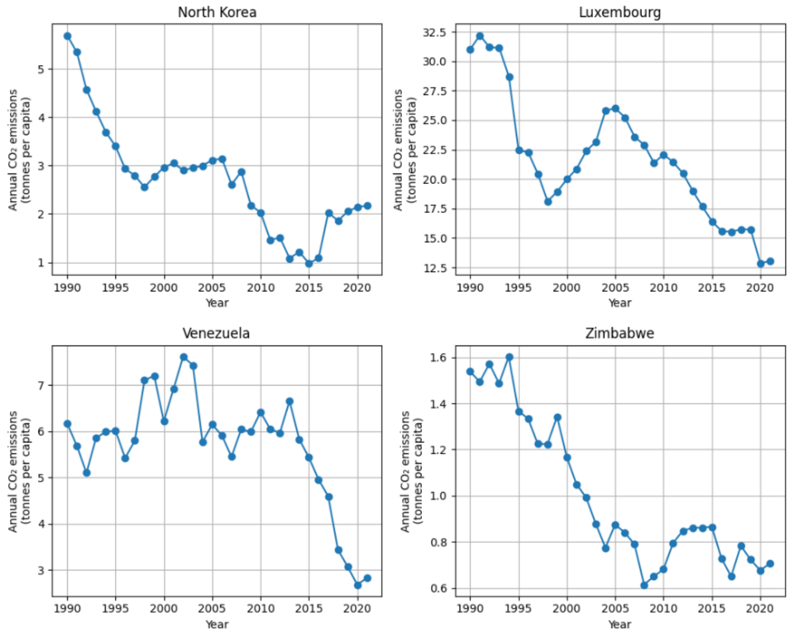
Functions for calculation correlation coefficient and p-values: https://www.statology.org/p-value-correlation-pandas/

Appendix

Plots showing the annual CO_2 emissions per capita of the top 10 countries of Q2a (Countries with the most substantial percentage decrease in CO_2 emissions per capita between 2020 and 1990):



Plots showing the annual CO_2 emissions per capita of the top 10 countries of Q2b (Countries with the most substantial percentage decrease in CO_2 emissions per capita between 2020 and 1990, with the condition that the CO_2 emission of the country >= 5 megatonnes). Only the countries are shown, which are not displayed in the former plots:



Plots showing the annual CO_2 emissions per capita of all the top 3 countries of Q2c (Recent Reductions (last 1, 2, 3, 5, and 10 years) in CO_2 emissions per capita relative to 2021):