FINAL ASSIGNMENT REPORT WINC ACADEMY
Tristan Schoen (20 March 2024)

1. Introduction

In assignment of Winc Academy, a data-driven attempt to answer a few questions using data analysis with Python was carried out. These questions are part of a final assignment, but they also shed light on important matters related to CO_2 emissions and green energy. The data used to analyze and produce graphs in this report is exclusively sourced from *ourworldindata.org*, and the respective sources can be found there. For CO_2-related questions the period 2000 – 2022 was analyzed, this is because:

- The Intergovernmental Panel on Climate Change (IPCC) in 1988 and the adoption of the United Nations Framework Convention on Climate Change (UNFCCC) in 1992 were only the beginning of widespread international awareness and measurement of CO_2-emissions.
- In 2001, the Greenhouse Gas Protocol was published, after a decade of development. It established standards and rules for the calculation of carbon emissions according to their scopes, and thus uniformity and accuracy in measurements.
- *ourworldindata.org* does not provide data after 2022.

The three questions that were analyzed were:

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?

A shortened summary of the results for each question, in respective order, is:

1. Among the analyzed predictors, GDP per capita, fossil fuel production and use of fossil fuel generated electricity all show a significant correlation with CO_2 emissions per capita. However, fossil fuel electricity used per capita shows the strongest correlation with CO_2 emissions per capita. (r = 0.8672)
2. Countries that show the biggest decrease in total CO_2 output as well as CO_2 output per capita are: Syria, Denmark, Venezuela, United Kingdom and Nauru. The individual causes or efforts of these decreases is left unanswered.
3. A model based on historical prices of green energy sources predicted that the cheapest green energy source in 2030 is photovoltaic solar energy, followed by onshore wind and concentrated solar power. For 2040 the model predicted photovoltaic solar energy, followed by concentrated solar power and onshore wind to be the cheapest sources of green energy.

2. Analysis

2.1 CO_2 Output Per Capita Predictor

In this section the data was used of the CO_2 output of 196 countries between 2000 and 2022 and several factors that are generally known to cause elevated CO_2 emissions. These include: GDP per capita, meat production, share of industry in GDP, fossil fuel production and use. This data will be plotted in line graphs to visualize the data. An organized overview of correlations and conclusions based on Pearson's correlation coefficient and p-value can be found in section 2.1.6.

A p-value of 0 indicates that the observed result is extremely unlikely to have occurred by random chance alone. In statistical hypothesis testing, the p-value represents the probability of obtaining a test statistic at least as extreme as the one observed, assuming that the null hypothesis is true (Rumsey, How to Find P Value from a Test Statistic, 2023).

A correlation coefficient (r) bigger than 0.7 indicates a strong linear relationship between the analyzed datasets, meaning that as one variable increases, the other variable tends to increase as well. (Rumsey, What Is R Value Correlation?, 2023)

2.1.1 GDP per capita

GDP per capita can be used as an expression of general welfare in a country. There are a lot of factors that influence and are influenced by what is considered "welfare". One example of the vast amount of factors is an increased amount of vehicles per capita as GDP per capita increases (Eder, Filimonova, Nemov, Komarova, & Sablin, 2019). The results of plotting GDP per capita against CO_2 emissions per capita is displayed in figure 1 (page 2).

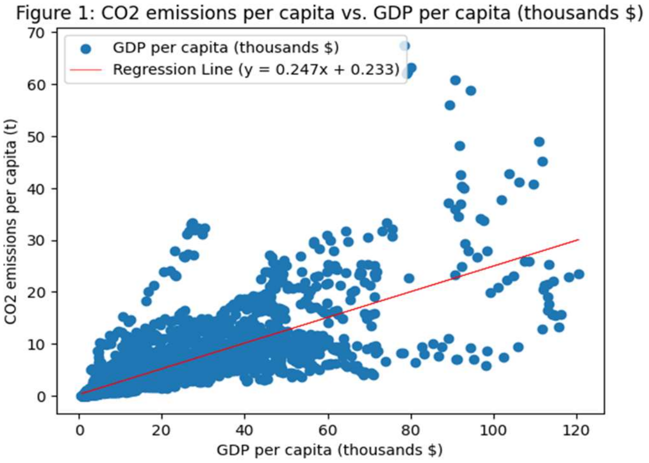

Figure 1: GDP per capita versus CO_2 emissions per capita of countries between 2000 and 2022. The red line is an applied linear trendline in an attempt to describe the correlation mathematically.

In figure 1 it is already visible that there seems to be a correlation between GDP per capita and CO_2 emissions per capita. The calculated Pearson's correlation coefficient 0.7646 with a P-value of 0.0 suggest that the data is highly correlated.

2.1.2 Meat production

Meat production is known to be the biggest source of producing CO_2 when it comes to food production. In food production, 60% of CO_2 emissions correspond to animal-based food, including meat. (Xu, 2021) The results of plotting GDP per capita against CO_2 emissions per capita is displayed in figure 2.

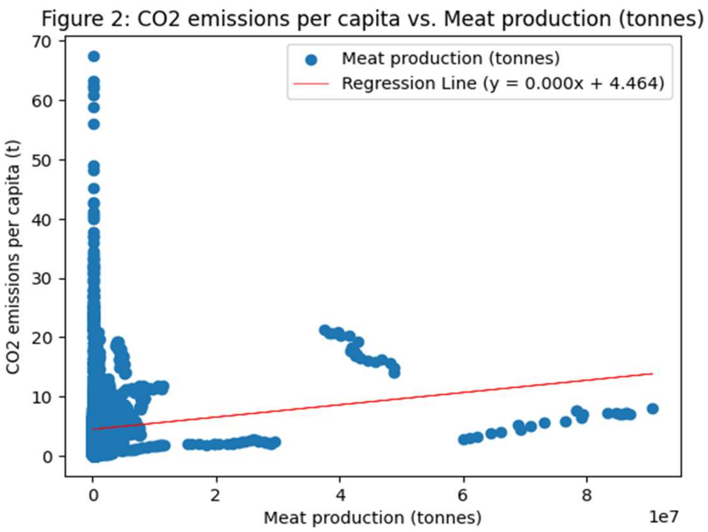

Figure 2: Meat production versus CO_2 emissions per capita of countries between 2000 and 2022. The red line is an applied linear trendline in an attempt to describe the correlation mathematically.

In figure 2 it is visible that a correlation between meat production and CO_2 emissions per capita is not obvious. The calculated p-value is 0.000. A low correlation coefficient combined with a very low p-value (such as 0.0) typically indicates that there is a statistically significant relationship between the variables being studied, despite the strength of the relationship being weak. The calculated Pearson's correlation coefficient 0.1102 suggests there is a negligible correlation.

2.1.3 Share of industry in GDP

Since the industrial revolution, CO_2 levels are on the rise, even until today. (National Oceanic and Atmospheric Administration, 2022). This raises the question if the share of the industrial sector in a country's GDP can be directly correlated to CO_2 emissions. The results of plotting the share of industry in GDP against CO_2 emissions per capita is displayed in figure 3.

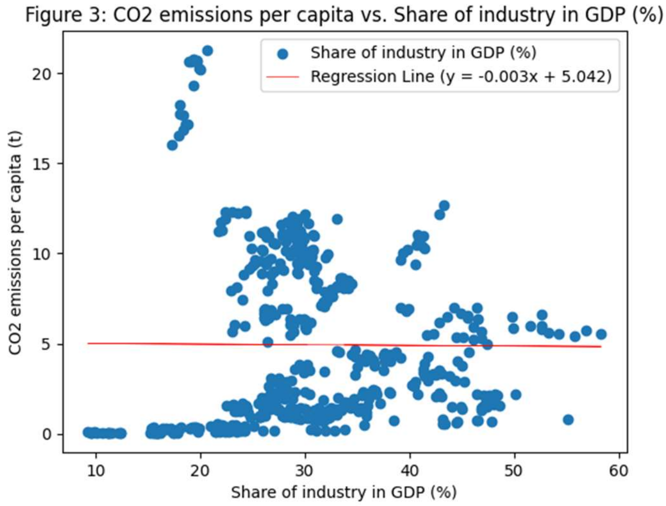

Figure 3: Percentual share of industry in GDP versus CO_2 emissions per capita of countries between 2000 and 2022. The red line is an applied linear trendline in an attempt to describe the correlation mathematically.

In figure 3 it is visible that a correlation between meat production and CO_2 emissions visually seems to be absent. The calculated p-value of the plotted data is 0.8766. A p-value close to $p = 1.0$ typically indicates there is no statistically significant relationship between the variables being studied. The calculated Pearson's correlation coefficient -0.0069 ($r \approx 0.0$) suggests there is no correlation between share of industry in GDP against CO_2 emissions per capita.

2.1.4 Fossil fuels produced per capita

When studying the dataset concerning CO_2 emissions per capita, some countries known as petrostates (i.e. Qatar, Kuwait, etc.) catch the eye in the top 10 of CO_2 emitters per capita. These countries are known for their intensive production of fossil fuels. This raises the question if the production of fossil fuels (coal, gas and oil: expressed in the energy-unit kWh) correlates with CO_2 emissions per capita. This relation is plotted in figure 4 (page 3).

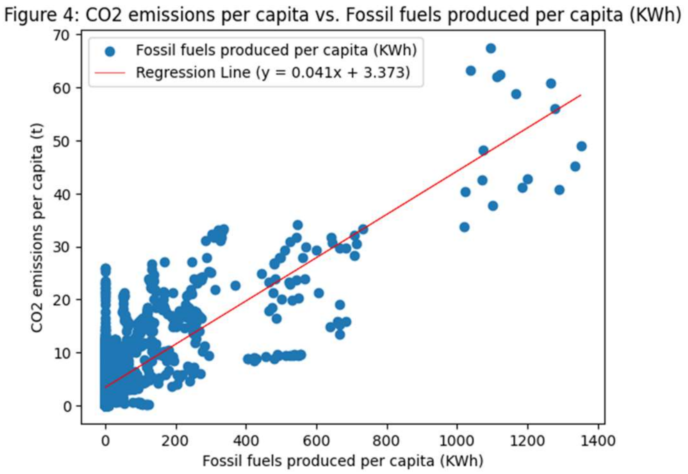

Figure 4: Fossil fuels produced per capita (includes: coal, gas and oil) versus CO_2 emissions per capita of countries between 2000 and 2022. The red line is an applied linear trendline in an attempt to describe the correlation mathematically.

In figure 4 it is already visible that there seems to be a correlation between fossil fuels produced per capita and CO_2 emissions per capita. The calculated Pearson's correlation coefficient 0.7464 with a p-value of 0.000 suggest that the data is highly correlated.

2.1.5 Fossil fuel electricity used per capita

Energy use and the energy sector is the biggest producer of CO_2 worldwide (Ritchie, 2020). Therefore a possible strong correlation could exist between CO_2 emissions per capita and the use of fossil fuel generated electricity per capita of a country. The use of fossil fuel generated electricity can be directly attributed to the intensity of (local) generation of electricity with fossil fuels and thus to CO_2 emissions per capita in a country. In figure 5 the relation between CO_2 emissions per capita and the use of fossil fuel generated electricity (expressed in kWh) is displayed.

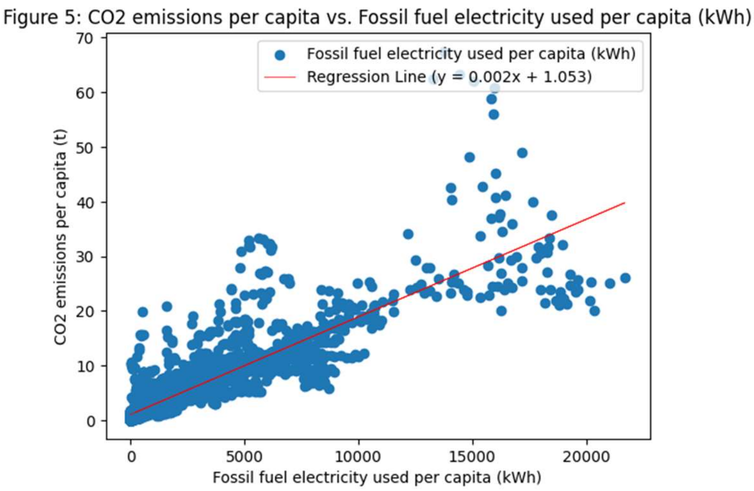

Figure 5: Use of fossil fuel generated electricity per capita versus CO_2 emissions per capita of countries between 2000 and 2022. The red line is an applied linear trendline in an attempt to describe the correlation mathematically.

In figure 5 it is already visible that there seems to be a correlation between the use of fossil fuel generated electricity per capita and CO_2 emissions per capita. The calculated Pearson's correlation coefficient 0.8672 with a p-value of 0.000 suggest that the data is highly correlated.

2.1.6 Conclusion

The general overview of the analyzed factors in regard to the emission of CO_2 per capita is displayed in table 1. For each factor it contains the correlation coefficient and the p-value to evaluate the statistical significance and the strength of the correlation with CO_2 emission per capita. It also contains the linear regression equation and the correlation of that respective trendline. This is done to determine how good a linear model fits the data analyzed.

Table 1: Overview of factors that were investigated versus the emission of CO_2 per capita of 196 countries between 2000 and 2022 and their respective correlation.

Factor versus emission of CO_2 per capita	Pearson Correlation Coefficient (r)	P-value	Linear Regression Equation	Correlation of Linear Trendline (r_t)
GDP per capita (fig 1)	0.765	0.000	y = 0.247x + 0.233	0.585
Meat prod. (fig 2)	0.110	0.000	y = 0.000x + 4.464	0.012
Share of industry in GDP (fig 3)	-0.007	0.877	y = -0.003x + 5.042	0.000
Fossil fuels produced per capita (fig 4)	0.746	0.000	y = 0.041x + 3.373	0.557
Fossil fuel electricity used per capita (fig 5)	0.867	0.000	y = 0.002x + 1.053	0.752

When looking at the correlation coefficients and p-value, it is visible that GDP per capita, fossil fuel production per capita and the use of fossil fuel generated electricity per capita of a country all highly correlate with CO_2 emissions per capita. (p-value = 0.0 and r > 0.7)

The strongest correlation is seen between CO_2 emissions per capita and fossil fuel electricity used per capita (figure 5, r = 0.867). This seems logical as the use of fossil fuels to generate electricity is one of the most direct ways to produce CO_2 emissions. Even a linear trendline trying to describe the data highly correlates with the data ($r_t > 0.7$).

GDP per capita as well as fossil fuels produced per capita show a high correlation with CO_2 emissions per capita. It is worth noticing that GDP per capita is a factor with an innumerable amount of causes and effects, but does not directly cause CO_2 emissions. It does not seem illogical that a lot of factors that bring prosperity to a country (and the results of this prosperity) are intensive CO_2 emitters. In respect to fossil fuels produced per capita: the production of fossil fuels possibly is not something that emits a lot of CO_2, but is a lucrative practice that influences the GDP per capita of a country. This shows how these factors can correlate with CO_2 emissions per capita, but perhaps are not the direct cause.

The share of industry in GDP does not seem to correlate with CO_2 emissions per capita (high p-value, $r \approx 0.0$). Possible causes can be that extensive industry is concentrated in a limited amount of countries, the variety between CO_2 emissions per type of industry or the impact of industry is too small compared to other sectors within a country.

The total meat production of a country seems to have a negligible correlation with CO_2 emissions per capita of the respective country. (p-value = 0.0, $r \approx 0.1$). This could be probably be attributed to the small impact this type of agriculture has in comparison to numerous other sectors in a country. Meat also is a heavily exported product by some meat producing giants such as the USA, Brazil and Australia (Workman, 2023). It possibly can be more favorable for a country to import than to produce meat on its own; this theory can be supported by the vertical concentration of points near the y-axis in figure 2, where countries have a negligible amount of meat production but a very variable CO_2 output per capita.

2.2 Biggest strides in decreasing CO_2 output

Between 1996 and 2010 international attention for global warming peaked, together with the subject of CO_2 emissions and its properties as a greenhouse gas. (Schmidt, Ivanova, & Schäfer, 2013). Since this surge in climatological awareness, various efforts have been done to reduce CO_2 emissions. Additionally, international agreements like the Paris Agreement (2016) have facilitated cooperation among nations to collectively address climate change through emission reduction commitments.

In this section data analysis has been performed on 196 different countries to compare their CO_2 emissions in 2000 and 2022 to determine which countries have the highest percentual decrease in CO_2 emissions between these dates. This is done for their total CO_2 emissions as well for CO_2 emissions per capita to see which countries made the biggest strides in decreasing CO_2 output.

2.2.1 Analysis

In figure 6 the top 10 of countries with the biggest decrease in total CO_2 emissions between 2000 and 2022 is displayed with their respective amounts of percentual decrease.

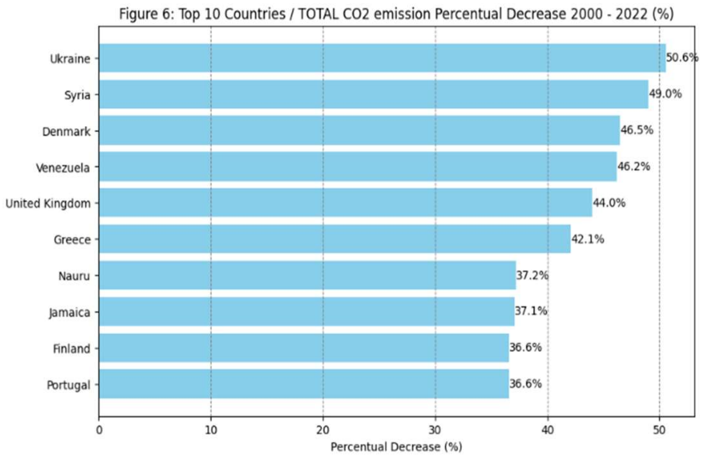

Figure 6: Top 10 of Countries with the biggest percentual decrease in total CO_2 emissions between 2000 and 2022.

The percentual decrease displayed in figure 6 is the absolute decrease in CO_2 output per country. These decreases are seen as a positive development, however it possibly does not highlight individual strides as countries differ in population. A sparsely populated country might halve its CO_2 output but compared to a highly populated country that has a slight decrease in CO_2 output it still might look minimal. Moreover, the population of a country can change; therefore a better way is to consider the change in CO_2 emissions per capita. This gives a more clear image of how intensive the CO_2 output is for certain countries. The top 10 of countries with the biggest decrease in CO_2 emissions per capita between 2000 and 2022 is displayed in figure 7.

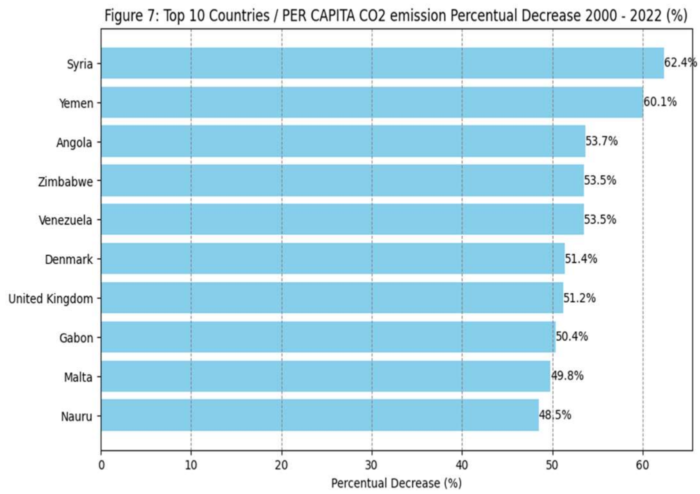

Figure 7: Top 10 of Countries with the biggest percentual decrease in CO_2 emissions per capita between 2000 and 2022.

When comparing figure 7 with figure 6 it is clear that both top 10 lists are different; there are countries that appear in both lists however. This means that these countries managed to decrease their total CO_2 output as well as CO_2 emissions per capita. Next, the conclusion section will provide more elaboration.

2.2.2 Conclusion

Table 2 shows countries that appear in both top 10 lists.

Table 2: Countries that appear in the top 10 of biggest percentual decrease in CO_2 emissions per capita as well as the top 10 biggest total percentual decrease in CO_2 between 2000 and 2022.

Country (2000-2022)	TOTAL CO_2 Decrease (%)	PER CAPITA CO_2 Decrease (%)
Syria	49.0	62.4
Venezuela	46.2	53.5
Denmark	46.5	51.4
United Kingdom	44.0	51.2
Nauru	37.2	48.5

The numbers in table 2 can be regarded as positive developments in the stride to decrease global CO_2 output. However the cause behind these decreases might differ per country. For example, the numbers for Syria and Venezuela might look positive but are likely the result of a crippled society and/or economy caused by unfortunate events in the past decades in these countries. The other three countries are not in these situations and are more likely to have decreased their CO_2 output with (innovative) efforts.

2.3 Non-fossil fuel energy price predictions

In the race to decrease CO_2 output, it is interesting to know what non-fossil fuels will be at a favorable price in the future. Predictions can be done with models based on historical prices of energy sources. In this section models are made with, among others, Scikit-learn features based on the historical prices of energy sources that are considered 'green'.

A problem that arises when using a linear model is that an energy source that rapidly falls in price, will eventually result in a negative price, which is unlikely. To ensure that energy costs never reach exactly zero, a positive value very close to 0 was added (epsilon) to energy costs and were changed using the natural logarithm (np.log(energy costs + epsilon)). After that, a simple polynomial model with a straight trendline (degree 1) was created and trained using the transformed data and then predictions were back-transformed to the original scale for interpretation (np.exp(predictions_transformed)). This model was used to predict energy costs for different sources.

As with all model predictions, only predictions in the near future can be regarded as (somewhat) reliable. In this case it was attempted to predict the prices in 2030 and 2040 with the model.

2.3.1 Analysis

In figure 8 the historical data is displayed of the price of the six most prevalent green energy sources in different colored (in $ / kWh) together with their models in red dotted lines. The prices used are the global average prices of the respective energy source.

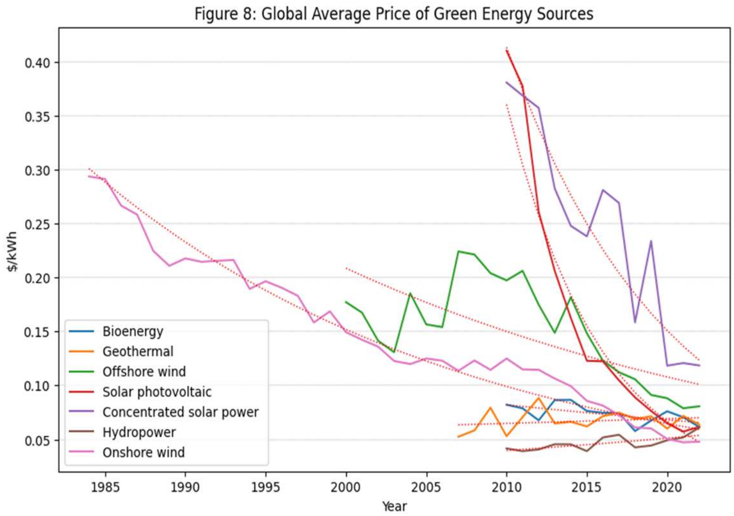

Figure 8: Historical data of the average price of green energy sources globally. The red dotted line is a prediction model for every energy type, to predict prices in 2030 and 2040.

As seen in figure 8 and as mentioned before, the prices of some energy sources have fallen drastically. In this case solar energy takes a price fall through recent years, but seems to make a curve to a less steep decrease as it gets near the value of 0.05 $/kWh. Thus a strictly linear model already will predict negative prices in the near future. The model was used to predict the cheapest green energy sources in 2030 and 2040. These predictions are respectively displayed in figure 9 and 10.

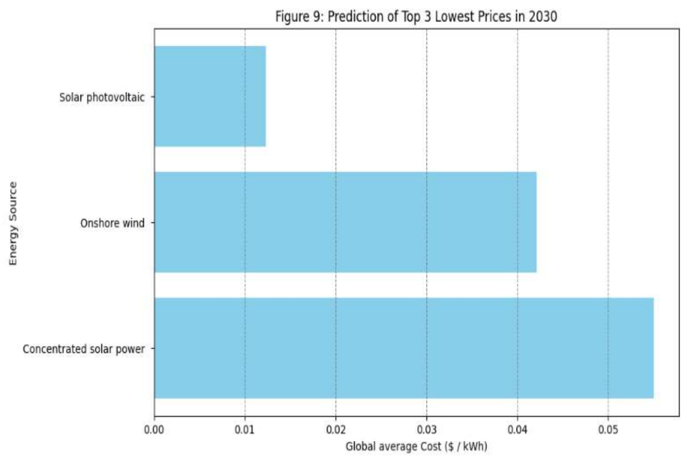

Figure 9: Top 3 lowest 'green energy' prices in 2030 predicted by a model based on historical prices of the 6 most prevalent green energy sources.

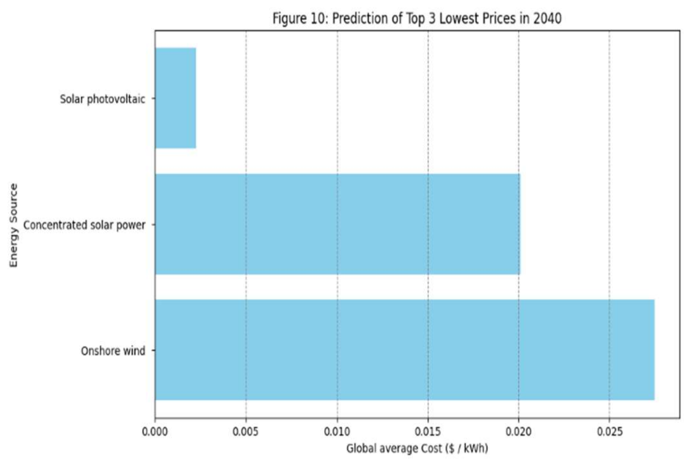

Figure 10: Top 3 lowest 'green energy' prices in 2040 predicted by a model based on historical prices of the 6 most prevalent green energy sources.

As seen in figure 9 and 10, the model predicted solar photovoltaic energy to be the cheapest green energy source for both 2030 and 2040. In 2040 concentrated solar power replaces onshore wind on the second place compared to 2030. (To clarify the differences; solar photovoltaic energy is energy directly generated from sunlight. Concentrated solar power is energy generated with heat produced by sunlight).

2.3.2 Conclusion

In table 3 an oversight of the top 3 cheapest green energy sources predicted for 2030 and 2040 by a model based on historic prices. As seen, in both years solar photovoltaic energy is predicted to be the cheapest source of energy. Onshore wind energy and concentrated solar power remain in the top 3 from 2030 to 2040.

Table 3: Predicted top 3 cheapest green energy sources for 2030 and 2024, based on a model trained with historic data on the global average of the six most prevalent green energy sources.

Year	Energy Source	Predicted Cost ($/kWh)
2030	Solar photovoltaic	0.0123
2030	Onshore wind	0.0421
2030	Concentrated solar power	0.0551
2040	Solar photovoltaic	0.0023
2040	Concentrated solar power	0.0201
2040	Onshore wind	0.0275

As a model is a means to predict future values instead of determining them. Models are something that always should be handled with a critical view, especially when these models try to describe data that is heavily influenced by external factors. Also, predictions become increasingly reliable as they approach the near future instead of more distant moments in time. In the case of the analyzed data for example (see table 3), the predicted cost price should not been seen as an absolute. A better way to view the data is to look at the relative position of the energy sources within the top 3 and in contrast to the other analyzed energy sources.
This shows that solar photovoltaic energy is a promising candidate to have the best price per kWh in the future when considering non-fossil fuel green energy sources, followed by concentrated solar power and onshore wind.

3. Conclusion / Discussion

The following questions have been answered in this report:

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?

In the following subsections these will be set out.

3.1 CO_2 Output Per Capita Predictor: conclusion / discussion

The analyzed predictors included GDP per capita, meat production, share of industry in GDP, fossil fuel production per capita and use of fossil fuel generated electricity per capita.

Among the analyzed predictors, GDP per capita, fossil fuel production and use of fossil fuel generated electricity all show a significant correlation with CO_2 emissions per capita. However, fossil fuel generated electricity used per capita shows the strongest correlation with CO_2 emissions per capita. ($r = 0.867$) followed by GDP per capita ($r = 0.765$) and fossil fuel production per capita ($r = 0.746$) all with a p-value of 0.0, indicating an extremely strong statistical significance, suggesting that the observed correlation is highly unlikely to have occurred by chance.

While GDP per capita and the amount of fossil fuels produced per capita are related to how much CO_2 each person produces, they likely are not direct causes. On the other hand, using electricity generated from fossil fuels is likely a more direct reason for CO_2 emissions.

A suggestion for a future research might be the research of the influence of GDP per capita on the use of (fossil fuel generated) energy and which economic sectors within a country use the most energy.

3.2 Biggest strides in decreasing CO2 output: conclusion / discussion

The total CO_2 output and the CO_2 output per capita were compared between 2000 and 2022 for 196 different countries.

Countries that show the biggest decrease in total CO_2 output as well as CO_2 output per capita are: Syria (resp: 49.0% and 62.4%), Denmark (resp: 46.5% and 51.4%), Venezuela (resp: 46.2% and 53.5%), United Kingdom (resp: 44.0% and 51.2%) and Nauru (resp: 37.2% and 48.5%).

The individual causes or efforts of these decreases is left unanswered. A suggestion for a future analysis might be a comparison of these decreases with several socio-economic factors in the respective countries to determine what might be the possible cause of these decrease in CO_2 per capita

3.3 Non-fossil fuel energy price predictions: conclusion / discussion

A model based on historical prices of green energy sources was made within Python using Scikit-learn features. (including the following sources: Bioenergy, Geothermal, Solar photovoltaic , Concentrated solar power, Hydropower and Onshore wind). predicted that the cheapest green energy source in 2030 is photovoltaic solar energy, followed by onshore wind and concentrated solar power. For 2040 the model predicted photovoltaic solar energy, followed by concentrated solar power and onshore wind to be the cheapest sources of green energy.

However, it's essential to note that these projections rely on global average prices for each energy source, which may vary depending on specific regional factors and market conditions.

While the model considered various non-fossil fuel sources, it did not incorporate nuclear energy. Nevertheless, nuclear energy represents a significant potential alternative worth further investigation due to its low carbon emissions. However, it's important to acknowledge that nuclear energy can be a contentious solution, with debates surrounding safety, waste management, and proliferation risks.

4. Sources

Eder, L., Filimonova, I., Nemov, V., Komarova, A., & Sablin, K. (2019). Ecological aspects of economical development: issues of forecast greenhouse gas emissions in road transport in Europe and regions of Russia. *2018 International Conference on Renewable Energy and Environment Engineering* (p. 7). Novosibirsk State University, Russia: EDP Sciences, 2019.

National Oceanic and Atmospheric Administration. (2022, June 3). *Carbon dioxide now more than 50% higher than pre-industrial levels*. Opgehaald van noaa.gov: https://www.noaa.gov/news-release/carbon-dioxide-now-more-than-50-higher-than-pre-industrial-levels

Ritchie, H. (2020, September 18). *Sector by sector: where do global greenhouse gas emissions come from?* Opgehaald van Published online at OurWorldInData.org: https://ourworldindata.org/ghg-emissions-by-sector#article-citation

Rumsey, D. J. (2023). How to Find P Value from a Test Statistic. In D. J. Rumsey, *Statistics For Dummies.* John Wiley & Sons, Inc.

Rumsey, D. J. (2023). What Is R Value Correlation? In D. J. Rumsey, *Statistics For Dummies.* John Wiley & Sons, Inc.

Schmidt, A., Ivanova, A., & Schäfer, M. S. (2013, July 18). Media Attention for Climate Change around the World: A Comparative Analysis of newspaper coverage in 27 countries. *Global Environmental Change*, pp. Volume 23, Issue 5, Pages 1233-1248.

Schoen, T. R. (2024, March 20). *USED CODE FOR DATA ANALYSIS*. Opgehaald van GITHUB PUBLIC FILE: https://github.com/Heineken-H41/WincFinal/blob/main/WINCFINAL.ipynb

Workman, D. (2023). *World's Top Food Exports Special Data Report*. Opgehaald van World's Top Exports: https://www.worldstopexports.com/worlds-top-food-exports-special-data-report/

Xu, X. S.-S. (2021). Global greenhouse gas emissions from animal-based foods are twice those of plant-based foods. *Nature Food (Retrieved from https://doi.org/10.1038/s43016-021-00358-x)*, pp. 2(9), 724–732.