CO2 emissions

Final assignment Data Analysis with Python

15-05-2023

Lesty Munoz
WINC ACADEMY

1. Introduction

The increasing concentration of carbon dioxide (CO2) in the atmosphere is a significant concern for global warming and climate change. In this report, we will explore three questions related to CO2 emissions and non-fossil fuel energy technology. We will use data analysis techniques to answer these questions based on the provided dataset.

In relation to CO2 emission and the reduction of CO2 emission three questions have been formulated:

1. What is the biggest predictor of a large CO2 output per capita of a country?
2. Which countries are making the biggest strides in decreasing CO2 output?
3. Which non-fossil fuel energy technology will have the best price in the future?

For every analysis, the collected data is reviewed, recalculated, cleaned and visualised. The data and visualisation is explained on the next pages. Tables and figures are added to make this explanation visual.

Question 1 will be discussed in chapter 3, question 2 in chapter 4 and chapter 5 will handle the third question. The last chapter (6) will summarize the conclusions.

2. Data collection

For this analysis, data from the website www.ourworldindata.com is used. This website contains data about the large topics in the world. Our World in Data collects data free for analysts and other interested pares to use. To answer the questions asked, I have browsed this website to obtain relevant CSV files about different variables and used them to make an analysis.

3. What is the biggest predictor of a large CO2 output per capita of a country?

The objective of this analysis is to identify the biggest predictor of a large CO2 output per capita of a country. The analysis was conducted using data from "ourworldindata" and the results are based on correlation analysis. Calculating the correlation coefficients between CO2 outputs per capita and various factors.

A selection of possible biggest CO2 predictors has been made:

- ✓ CO2 output (file contains total CO2 output in tons per year, recalculated to kg/per year/per capita)
- ✓ Total population
- ✓ population density
- ✓ Share of urban population (file contains absolute numbers, recalculated to share of total population)
- ✓ Share of rural population (file contains absolute numbers, recalculated to share of total population)
- ✓ GDP and GNI per capita
- ✓ Gini coefficient
- ✓ Human development index (not recalculated, calculation based on life expectancy, GNI and education)
- ✓ Share of population under the poverty line
- ✓ Share of employees working in industry
- ✓ Use of different energies in kwh per capita (file contains use of energy in twh, recalculated to kwh)
- ✓ Agricultural output in US$ per capita
- ✓ Food supply in Kcal per capita per day
- ✓ Share of land area used for agricultural purpose

I used the following information about the degree of correlation from www.statisticssolutions.com:

- High degree: If the coefficient value lies between ± 0.50 and ± 1, then it is said to be a strong correlation.
- Moderate degree: If the value lies between ± 0.30 and ± 0.49, then it is said to be a medium correlation.
- Low degree: When the value lies below ± .29, then it is said to be a small correlation.

3.1 Results

The Pearson correlation coefficient has been calculated to find the relation between the CO2 output per capita of a country and possible predictors. It does not say what relation (causal or not), but in this case if there is no or only a small connection, the variable cannot be used.

	Variable	Correlation coefficient between variable and Annual CO_2 emissions (kg/py/pc)
0	Share of pop < poverty line	-0.448071
1	Rural (% of total)	-0.417690
2	Gini	-0.322609
3	Population	-0.049650
5	Population density	-0.038160
6	Industry (% of total employment)	0.053035
7	Agricultural land (% of land area)	0.078350
8	Agri output $	0.095496
9	Renewable kWh/pc	0.107329
10	Nuclear kWh/pc	0.214055
11	Urban (% of total)	0.417690
12	Food supply (kcal/pc/pd)	0.572206
13	HDI	0.574953
14	GDP pc	0.599668
15	GNI	0.643922
16	Fossil kWh/pc	0.960218

Table 1. The Pearson Correlation Coefficients of the annual CO2 emission per capita and the selected possible predictors.

To visualize the CO2 output compared to the variables in scatterplots with an added linear regression line (see figure 1 for the plots). These are the variables with a high correlation degree (above 0.5). Visualising the values makes it easier to see how the values are scattered and how the regression line relates to these values. The regression line shows how much and in what direction the response variable changes when the explanatory variable changes.

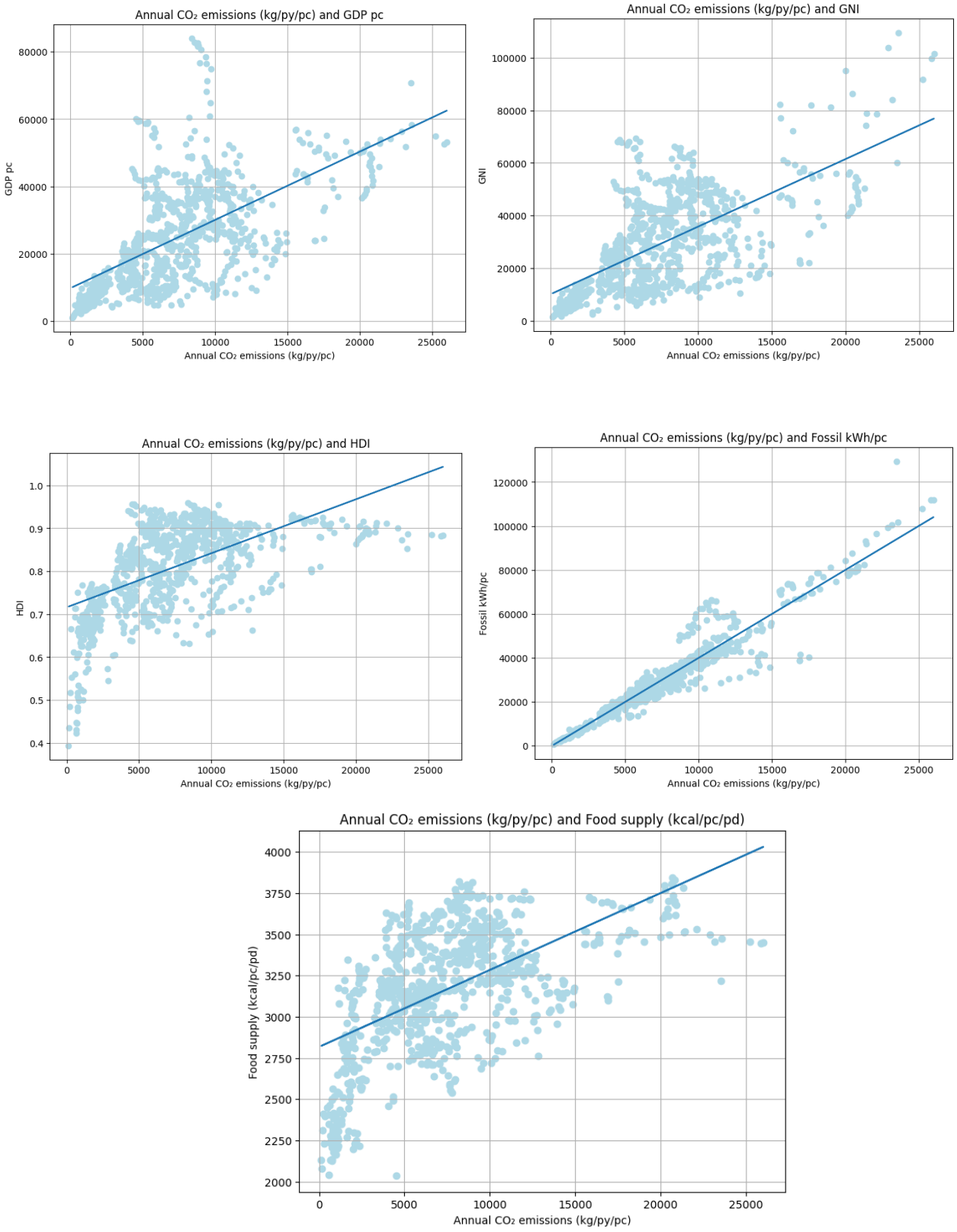

Figure 1. Scatter plot that shows the correlation between annual CO2 emission per capita and the different sources for fossil energy.

Summary of Findings:

The biggest predictor is fossil fuel consumption with a correlation coefficient of 0.96. The second biggest predictor is Gross National Income (GNI) with a correlation coefficient of 0.73.

There is a moderate to steep inclining slope on each regression line. Nevertheless, because the values are very much scattered or closed together in one place, the regression line does not really show a relation to the values. Only the variable for fossil fuel use in kWh per capita shows almost linear values without major outliers and a regression line that follows the values very close.

3.2 Conclusion

- The biggest predictor of a large CO2 output per capita of a country is the use of (total) fossil energy use.

- Only the variable for fossil fuel use in kWh per capita shows almost linear values without major outliers and a regression line that follows the values very close.

- This is the best variable to use to predict. In combination with the other 4 variables, I can conclude the following:

- To answer the question: What is the biggest predictor of a large CO2 output per capita of a country? A large use of fossil fuels per capita in combination with a great prosperity is the biggest predictor of a large CO2 output per capita of a country

4. Which countries are making the biggest strides in decreasing CO2 output?

To answer this question the change in annual CO2 emissions in the period 1961-2019 has been determined for countries of which data were available. The data is transformed to get a dataframe with the years set as columns and countries as rows. After dropping rows with missing values, the resulting dataframe contains values for 134 countries and 58 years.

The used variables are:

- CO2 output (file contains total CO2 output in tons per year, recalculated to kg/per year/per capita).

- Total population. The latest data that are available at this moment are from 2019.

4.1 Results

Table 2 shows the first measurement from 1961. This is point zero with a value of 100%. The next relative data points are calculated every 5 years with reference to 1961 and added in new columns. To make the dataframe smaller, the columns for the original years are dropped. The last measurement is from 2019 and is the endpoint. The calculated data is merged into a dataframe (see table 2), sorted from low to high (based on the values in column for relative CO2 output in 2019) and the best 10 are plotted in a line plot (see Figure 2).

Year	rlt1961	rlt1966	rlt1971	rlt1976	rlt1981	rlt1986	rlt1991	rlt1996	rlt2001	rlt2006	rlt2011	rlt2016	rlt2019
Country													
Zambia	100	80.381	74.704	68.684	49.425	36.096	26.235	17.922	16.059	16.177	20.519	30.322	36.586
United Kingdom	100	101.735	106.140	95.535	89.264	90.018	95.414	90.730	87.838	83.870	66.654	54.636	49.053
Venezuela	100	92.732	86.997	68.789	95.637	99.654	92.683	88.310	112.624	96.288	98.528	80.768	50.116
Zimbabwe	100	102.121	126.297	133.898	97.191	112.843	115.866	103.457	81.461	65.207	61.713	56.534	56.149
Sweden	100	143.212	161.647	165.241	128.694	113.845	104.135	111.661	96.899	91.400	80.591	67.364	61.640
Mozambique	100	72.661	110.151	67.088	56.020	19.909	18.403	17.860	22.434	24.300	34.391	77.490	62.071
France	100	114.202	141.488	145.223	124.296	103.047	107.948	101.806	101.941	97.519	82.561	76.389	71.805
Djibouti	100	246.101	212.256	175.207	178.747	180.620	92.334	92.332	97.196	98.390	102.453	75.437	74.231
Germany	100	109.481	116.478	122.728	118.999	118.698	112.008	103.965	99.090	95.351	87.590	85.706	74.982
Denmark	100	151.941	167.553	172.217	146.954	167.452	181.120	207.293	151.739	158.614	115.481	94.435	77.724

Table 2. Top 10 biggest decreasers in realtive CO2 emission since 1961.

Summary of Findings:

The analysis identified countries that exhibited substantial decreases in CO2 output over the specified period. These countries demonstrated a combination of effective climate policies, investments in renewable energy sources, and energy efficiency improvements. It is crucial to acknowledge that the analysis focused on relative improvements, considering the starting point and the overall progress made by each country.

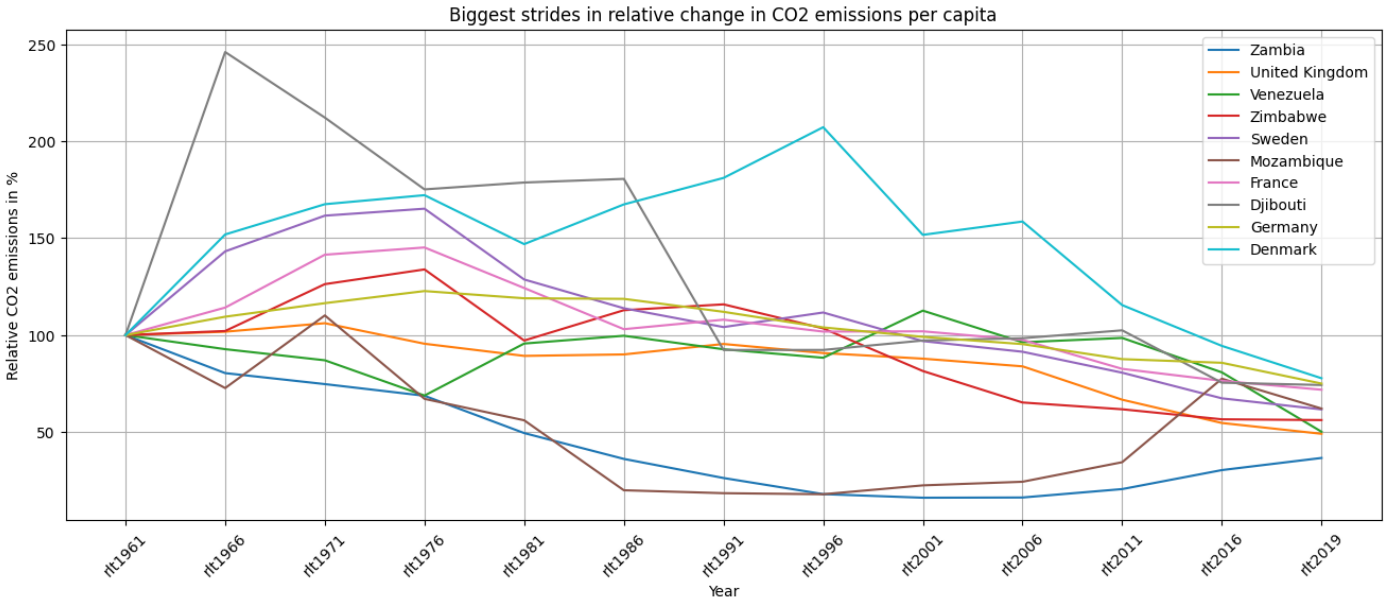

Figure 2. Line plot that shows the top 10 countries that made the biggest strides in decreasing CO2 output.

4.2 Conclusion

To answer the question: Which countries are making the biggest strides in decreasing CO2 output?

The top 10 of countries that made the biggest strides in decreasing CO2 output are (in order from greatest to least):

1. Zambia
2. United Kingdom
3. Venezuela
4. Zimbabwe
5. Sweden
6. Mozambique
7. France
8. Djibouti
9. Germany
10. Denmark

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

To answer this question, the levelized cost of electricity (LCOE) for various non-fossil fuel energy technologies was used. The LCOE represents the total cost of generating electricity over the lifetime of a plant divided by the total amount of electricity generated. The lower the LCOE, the cheaper the electricity generation. To predict which non-fossil energy technology will have the best price in the future a linear regression model has been fit, the missing values (NaN in the dataframe) are not replaced, because the course of the price development is not linear. The included variables prices of the following renewable energy sources are:

- Bioenergy
- Geothermal energy
- Solar photovoltaic power
- Concentrated solar power
- Offshore wind power
- Onshore wind power
- Hydropower

5.1. Results

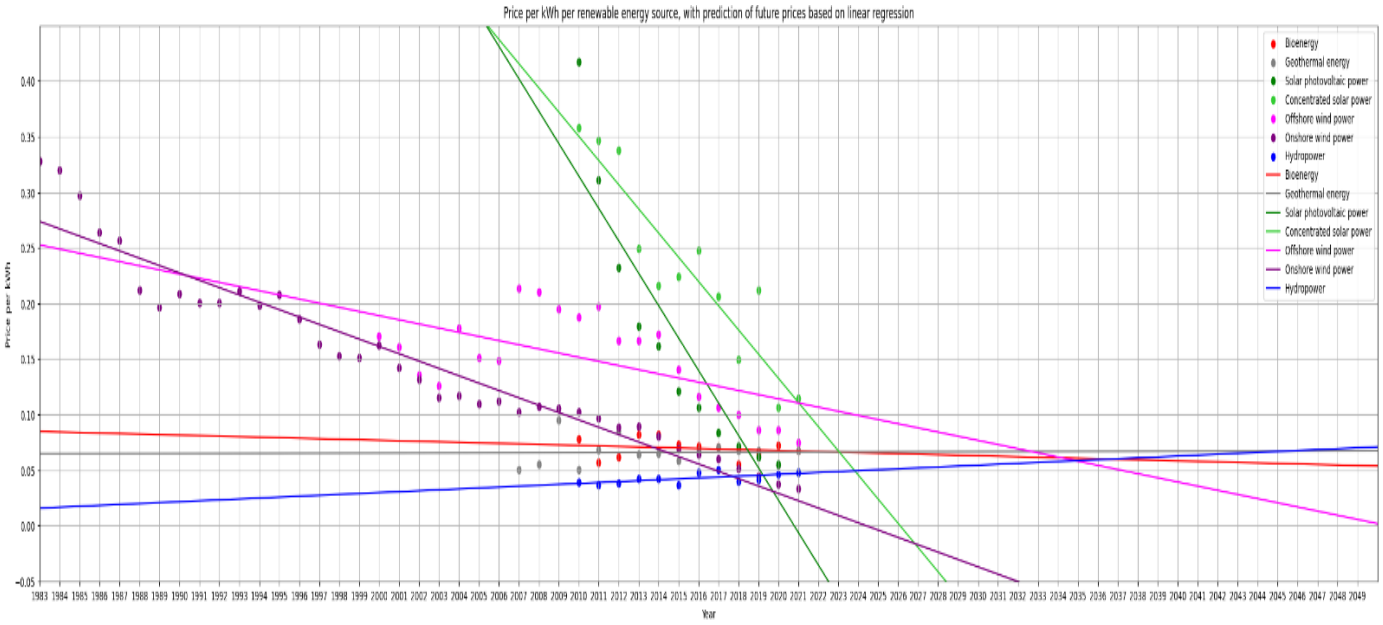

Figure 3. Levelized costs of energy by renewable energy sources with linear regression lines.

Summary of Findings:

The “levelized cost of energy by renewable energy source” graph shows the linear regression lines of all non-fossil energy sources. Although the prices of energy will not go below zero, the plot in figure three shows which energy source will have a lower price per kWh in the future. Bioenergy, geothermal energy and hydropower will roughly stay the same or increase in price. Solar photovoltaic power, concentrated solar power, offshore and onshore wind power will decrease in price.

5.2 Conclusion

Energy produced by solar panels has been decreasing fastest over the years, so it is most likely that solar power will have the best price in the nearest future (before 2030). Wind power will have a better price too, but a few years later.

APPENDICES

I. For detailed results, computer outputs and extra graphs of all the performed analyses of chapter 3 (What is the biggest predictor of a large CO2 output per capita of a country?), see:

https://github.com/LestyMV/Final-assignment---CO2-emissions/blob/main/Final_assignment_CO2_emissions_Part_1.ipynb

II. For detailed results and computer outputs of all the performed analyses of chapter 4 (Which countries are making the biggest strides in decreasing CO2 output?), see:

https://github.com/LestyMV/Final-assignment---CO2-emissions/blob/main/Final_assignment_CO2_emissions_Part_2.ipynb

III. For detailed results and computer outputs of all the performed analyses of chapter 5 (Which non-fossil fuel energy technology will have the best price in the future?), see:

https://github.com/LestyMV/Final-assignment---CO2-emissions/blob/main/Final_assignment_CO2_emissions_Part_3.ipynb