Project Assignment: CO2 Emissions

23.11.2024

To see the code, see: Eindopdracht Winc - CO2 Emissions

1. Introduction

In recent years the topic of climate change is rising. The primary reason for climate change is high CO2 emissions in the world. In this report we want to answer three questions related to CO2 and energy to get a better understanding of this topic.
The questions are:

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?

2. Data Cleaning

In order to answer our questions we used a total of three datasets from the website “Our World in Data”.
We merged the datasets about CO2 emissions and energy consumption to answer the first two questions. To get a compact dataset with only necessary data to answer the research questions, all extra columns were dropped. Based on a look at the raw data we dropped all rows that contained information before 1990 and after 2022 as there seemed to be a lot of missing values before and after that.
After that we concentrated on the missing values and used multiple imputation strategies to filter out irrelevant information and missing values. One important step here was to drop all the rows that didn’t have an ISO code, as they were not relevant to answer the questions and would have distorted the answers. Secondly we filled all empty values for the energy consumption columns with zero as we assume that that sort of energy is not available for the country.
The last important step in the cleaning phase was to take a look for possible outliers. One outlier was the CO2 output per capita from Kuwait in 1991. After research we discovered that this is probably due to the fact that their oil fields were burning because of the gulf war, therefore we excluded this data.
For question 1 we created a new column with values for gdp per capita as a possible predictor of a high CO2 emission and for question 2 we created a column for the relative change of CO2 emissions per capita. This way we took the developments within the population into account.
For question 3 we took a different dataset from the website “Our World in Data” and dropped all rows that weren’t data about the whole world.

3. Methods & Results

3.1. Question 1: What is the biggest predictor of a large CO2 output per capita of a country?

To answer this question we looked at the correlations between CO2 emissions per capita and possible predictors of a high CO2 emission. An overview of the correlations > 0.5 is displayed below in table 1.

Table 1 : Table representing the correlations between CO2 emission per capita of a country and possible predictors

Predictor	Pearson Correlation Coefficient CO2 per Capita
CO2 per capita	1.000000
Fossil energy per capita	0.917213
Energy per capita	0.854308
Gas energy per capita	0.833900
Oil production per capita	0.700306
Gdp per capita	0.597491
Oil energy per capita	0.590587

The table shows that there is a very strong correlation between consumption of fossil energy per capita and CO2 emission per capita. We can conclude that the biggest predictor of a high CO2 emission per capita is consumption of fossil energy per capita. Figure 1 shows the correlation in detail.

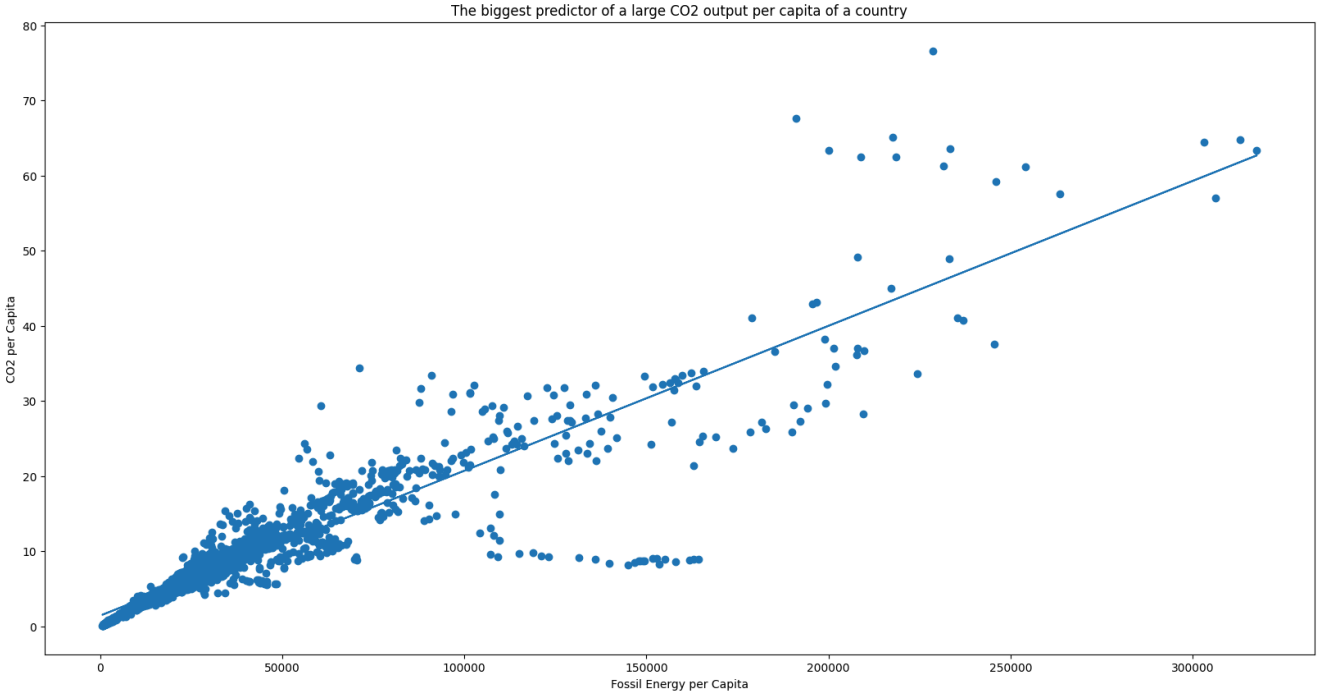

Figure 1: Figure shows the correlation between CO2 emission per capita and consumption of fossil energy per capita

3.2. Question 2: Which countries are making the biggest strides in decreasing CO2 output?

To answer which countries are making the biggest stride in decreasing their CO2 output we compared the CO2 emissions per capita of each year with our baseline, which was the CO2 emissions per capita of 1990. Based on that we took the top 5 countries with the largest relative decrease, which resulted in the outcomes shown in figure 2.

In conclusion we can say that Ukraine had the strongest decrease in CO2 output per capita since 1990.

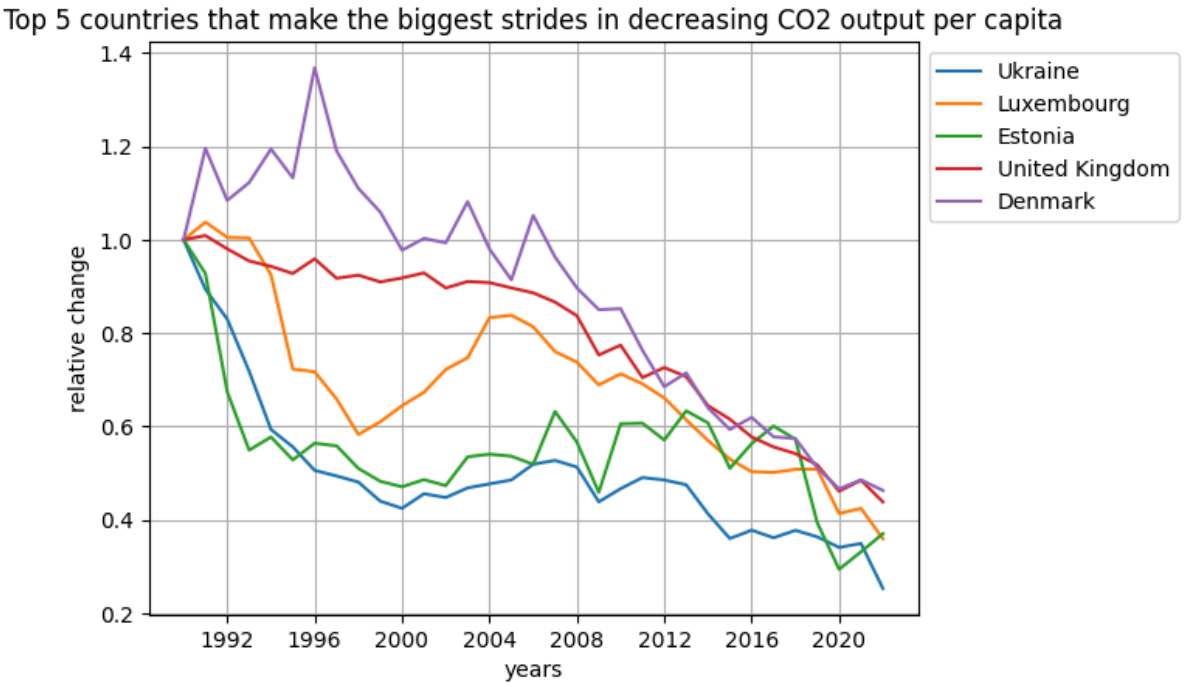

Figure 2: Figure shows the 5 countries that have decreased their CO2 emission based on a relative change since 1990

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

In order to answer which non-fossil fuel energy technology will have the best price in the future we used linear regression. This way we can show a trend line for the non-fossil fuel energy prices in the past and future.

Based on this we can conclude that solar photovoltaic energy will be the non-fossil fuel energy with the lowest price.

Figure 3 shows an overview of the different energy technologies and the trends on their prices.

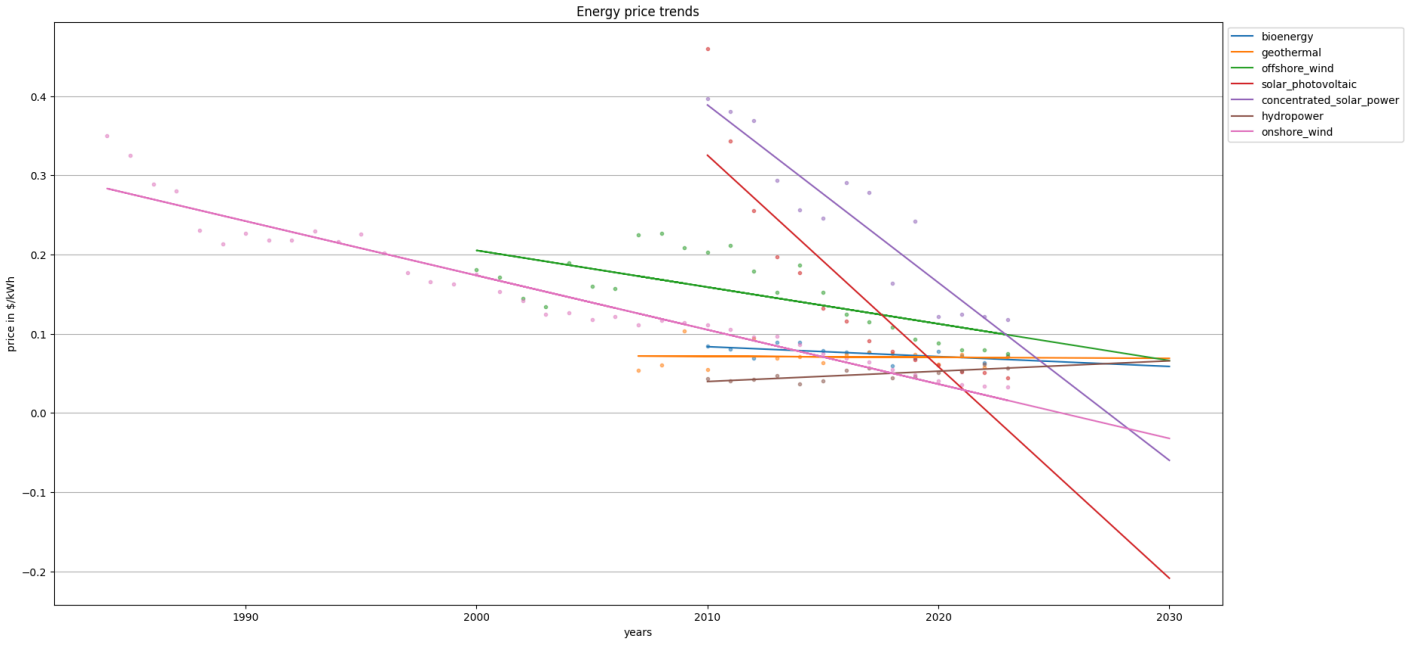

Figure 3: Figure shows energy prices over the years with a trendline which includes future years

4. Discussion & Limitations

This research offered some points of discussion and some limitations.
First we didn't gather the data or processed the original data but worked with datasets that were made by another party. Therefore we don’t know everything that happened prior to us using the data which could lead to extra biases.
Further points that could lead to biases are: we dropped a lot of data, like certain countries or years and we assumed certain data to be zero instead of not available. Both of these points could lead to biases or misformed information.
There is also room for bias, errors or different interpretations within our analyses of the questions.
One possibility in question 1 is that we didn’t think of the right predictors and therefore missed one that would have been an even better predictor of CO2 emission.
For question 2 we used a baseline of the year 1990 but the relative change per country would change if we would take a different year. For example, if we took the year 1996 as the baseline, Denmark would probably have a stronger relative change than e.g. Ukraine. Because of that the top 5 could even look completely differently. This is something to keep in mind.
Lastly we can see in figure 3 that three of the non-fossil fuel energy technology prices drop below zero according to our trend line. This is obviously not likely to happen, so linear regression might not be the best form of analysis for this.
Overall all these limitations should be kept in mind while looking at the results.