September 26, 2023

Final assignment – CO_2 emissions

Data Analytics with Python course at Winc Academy

Author: Monika Gostyńska

Assignment is to answer three following questions using any available data set from **Our World in Data**[1] website:

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?

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

Link to Notebook:
https://github.com/MonikaGost/Winc/blob/7e0d2c55186b6f4c770ec7c3d1d8b8171c95c4d1/Final_assignment/Final_assignment_Part1_CO2_predictors.ipynb

To be able to answer this question I started with looking for basic information about what are CO2 and CO2 output per capita actually, where do these come from and what is the data set based on. I assume most people know what is CO2 and where does it come from but I find it necessary, for the purpose of this analysis, to support these 'assumptions' with some data. This information will also help me to make choices in further analysis.
I have found lots of information on Our World in Data website and it is clear to me that CO2 emissions is quite a complex subject. I have made analyses of 10 data sets but found not all of them useful enough, so I decided not to include all of them in this report. These are to be found in my Jupyter notebook though. For the sake of clarity of this report I will try to keep it as simple as possible and do not go into too many details.

CO2 emissions data

Among available data on CO2 emissions I have found out that there are 2 main methods of measuring CO2 data per country. We have production-based emissions and consumption-based emissions. Emissions from international aviation and shipping are not included in any country or region's emissions. They are only included in the global total emissions. Also land use emissions are not taken into account for both methods.
Production- based emissions are measured from fossil fuel and industry emissions of a country (emissions produced domestically).
Consumption-based emissions are calculated by adjusting production-based emissions for trade (production-based emissions minus emissions embedded in exports, plus emissions embedded in imports). They attribute the emissions generated in the production of goods and services according to where they were consumed, rather than where they were produced. Therefore they reflect the consumption and lifestyle choices of a country's citizens. **Because of that I will use in most of the CO2 analysis per capita the data of consumption based emissions [2] rather than production based emissions.**

[1] https://ourworldindata.org
[2] Consumption-based emissions are not available for all countries because not all countries have sufficient, high-quality trade data. But those without complete data are a small fraction (3%) of the global total.

So where do CO2 emissions come from?

Looking for sources of CO2 I have found below plot (**Figure 1**). It shows distribution of total CO2 emissions among sectors worldwide. It says that energy sector is responsible for 73% of total emissions. I want to analyze available data on this subject further on.

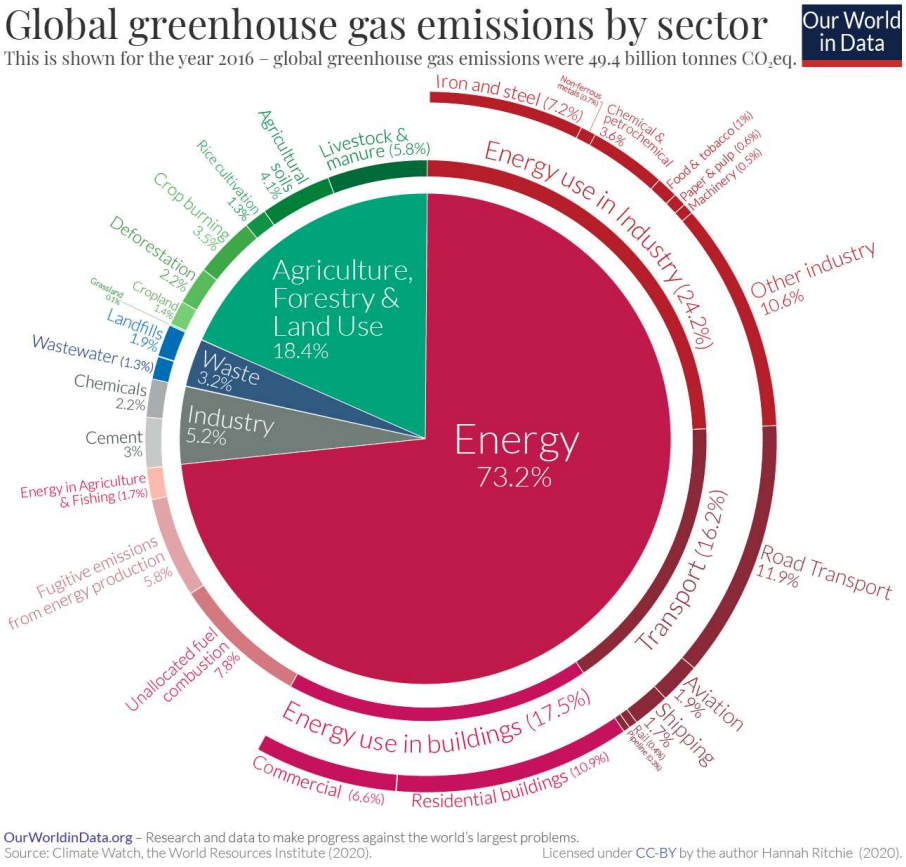


Figure 1 Distribution of global CO2 emissions by sector in 2016. Source: https://ourworldindata.org/emissions-by-sector

Similar, sector-related data on CO2 is found is the one below, measured per capita by sectors. It supports my thesis to look for CO2 drivers in energy sector (see. **Figure 2**).

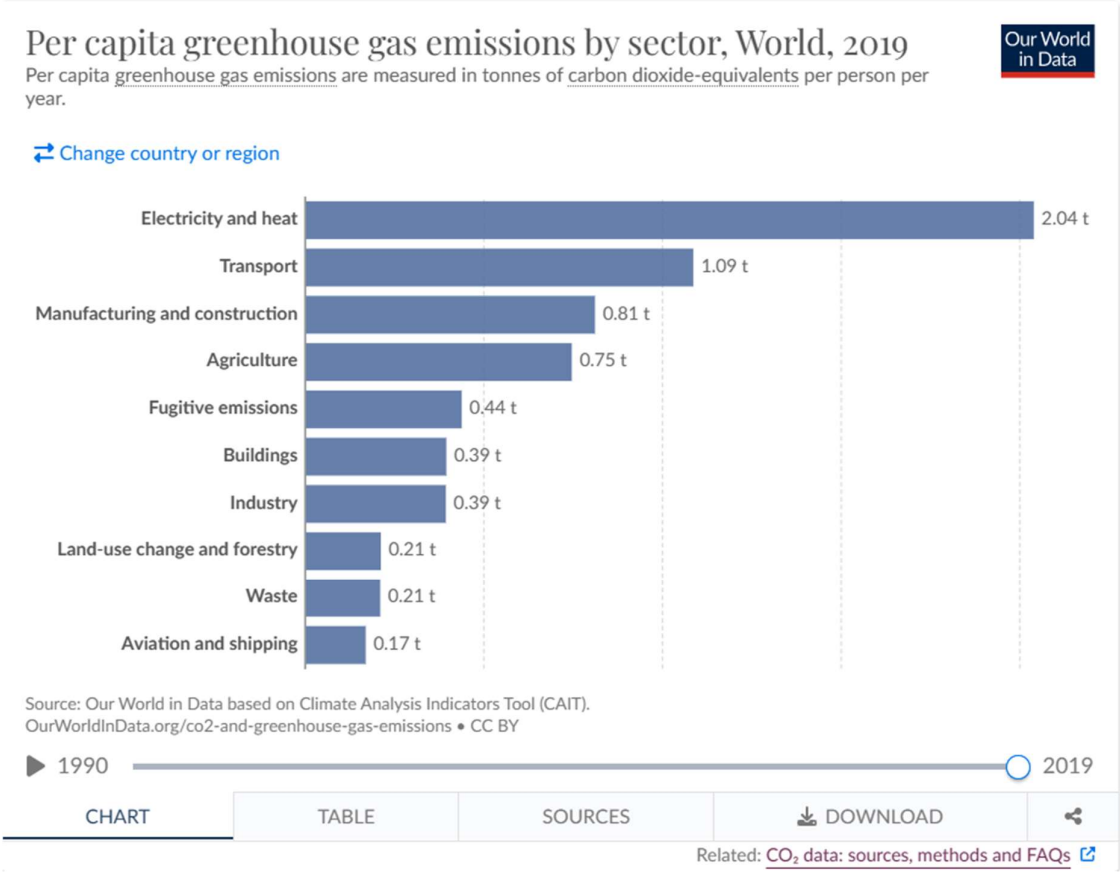


Figure 2. CO2 emission per capita by sector. World. 2019. Source: https://ourworldindata.org/emissions-by-sector

Per capita emissions represent the emissions of an average person in a country or region – they are total emissions divided by population. I will analyze these further in detail in the next paragraph.

Analyzing data sets

1. **Per capita consumption-based CO2 emissions (codes in Jupyter-notebook paragraph 2).**

Analysis of data of 120 countries, due to data availability. The most recent data is from the year 2020.

Looking at top 20 and bottom 20 countries with biggest CO2 emissions per capita per country we can draw the conclusion that the biggest CO2 emissions per capita belong to citizens from oil producing countries **(Figure 3)** and the lowest emissions are accountable to African countries **(Figure 4)**. This clearly suggests there may be a correlation with GDP, which I will investigate further on.

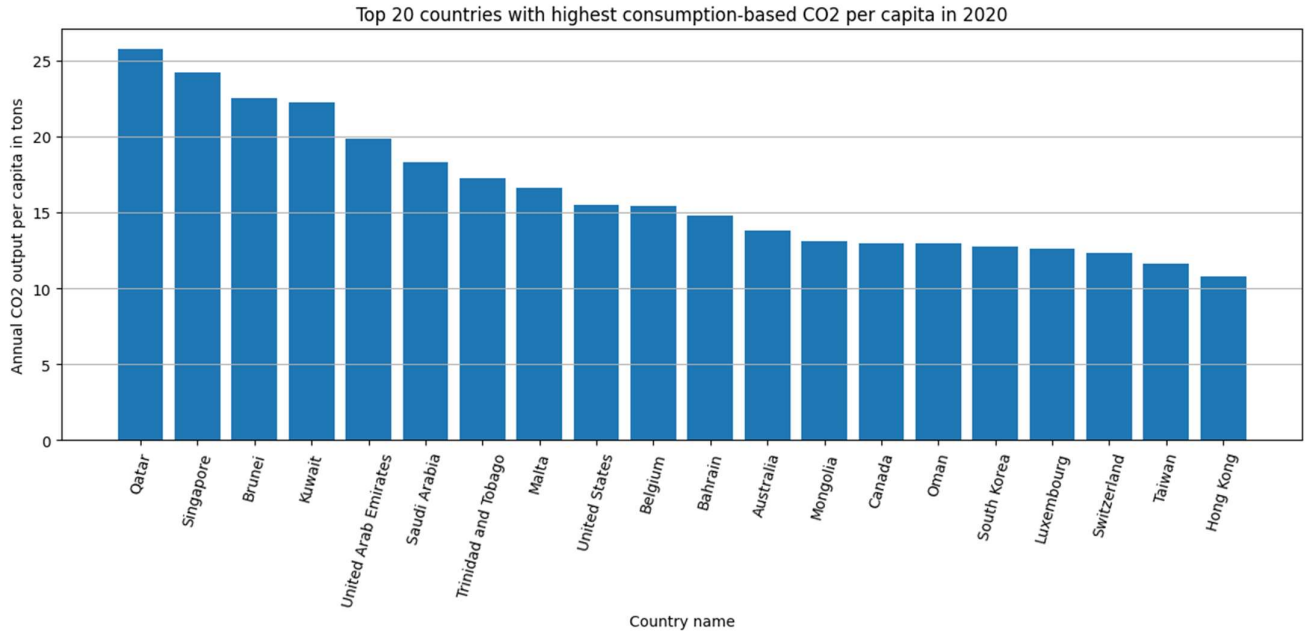


Figure 3.

Top 20 countries with highest CO2 output per capita in 2020.

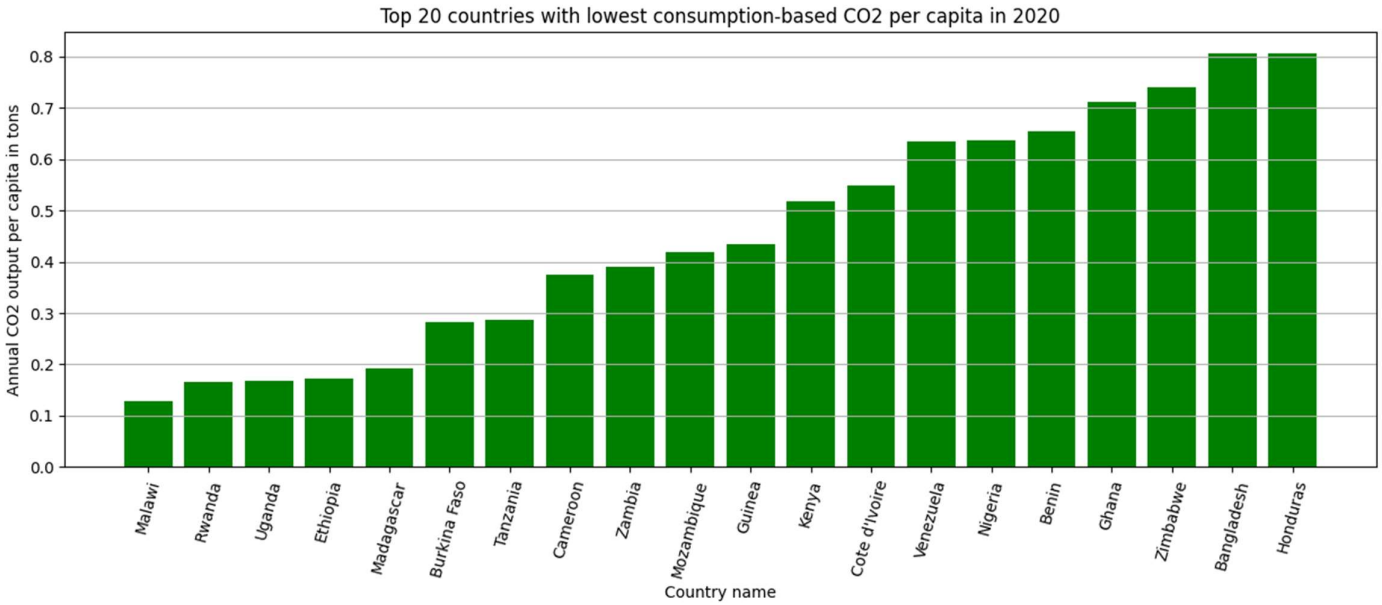


Figure 4.

List of 20 countries with the lowest CO2 output per capita in 2020.

Surprisingly, these are not the countries with the highest population number (**see Figure 5**) on the top of this list, nor the countries with largest global CO2 emissions **(Figure 6)**. Number of CO2 per capita will be higher if country happens to have large production-based CO2 emissions and relatively low population number. Like this relatively small population is accounted for large CO2 emissions output. In case of oil producing countries the CO2 output per capita is strongly affected by country's production activities where individuals have not personally a lot of impact on.

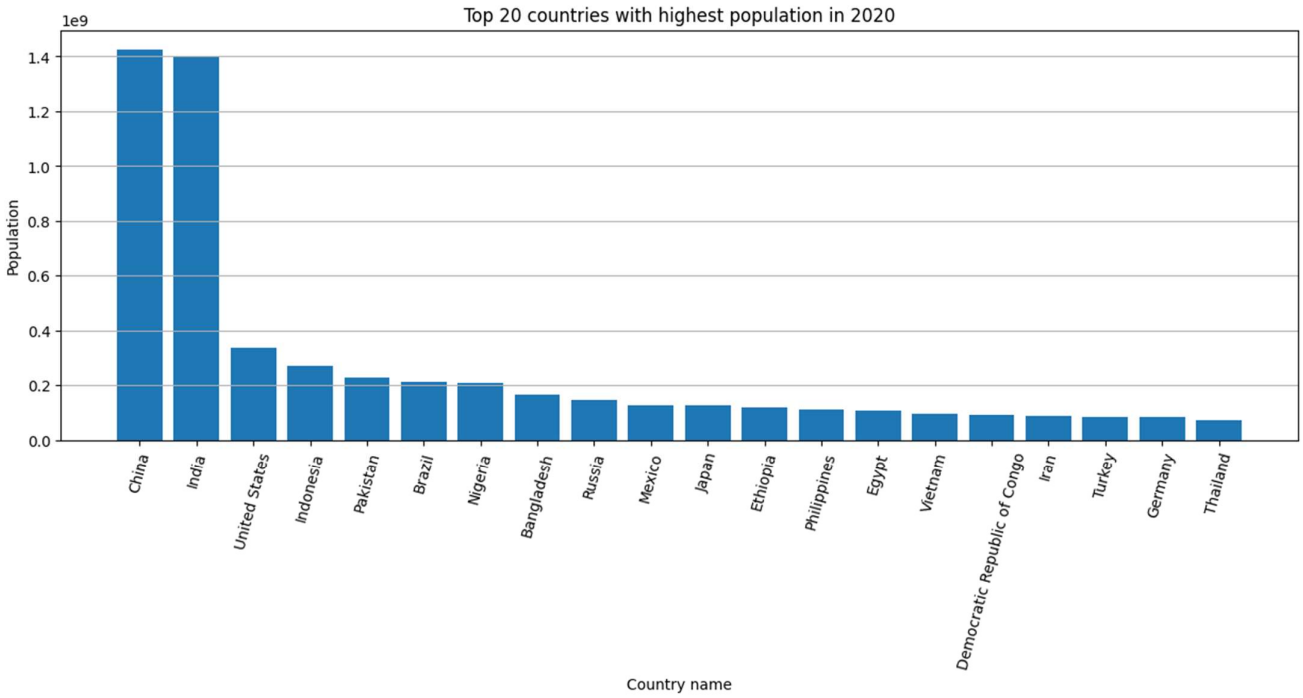


Figure 5. Top 20 countries with highest population number in 2020. (code in Jupyter-notebook paragraph 3).

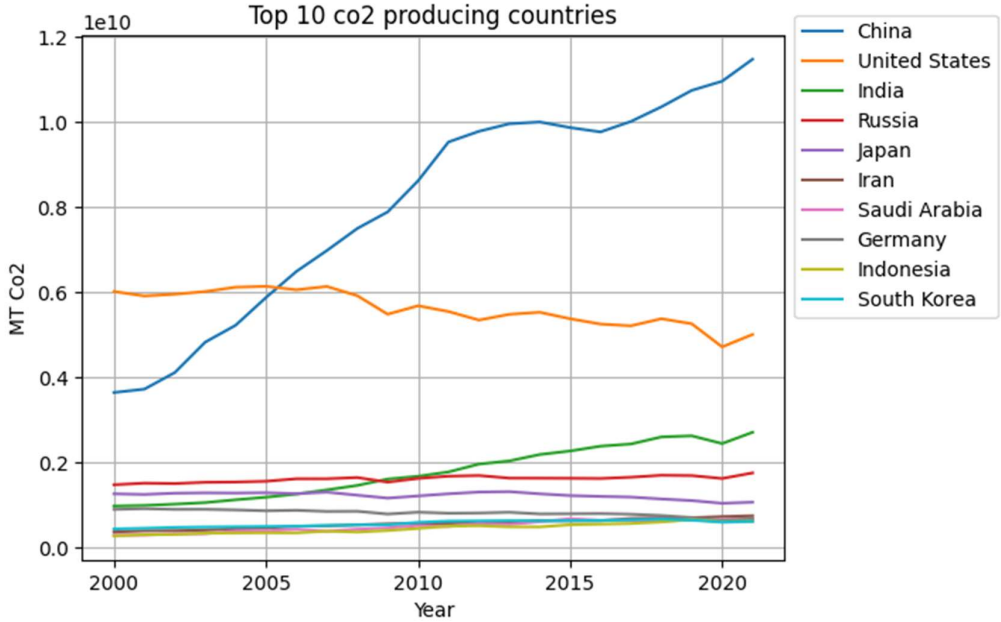


Figure 6. Top 10 Co2 producers. Data 2000-2020. (code in Jupyter-notebook paragraph 4).

2. **GDP per capita** (codes in Jupyter-notebook paragraph 6)

Data of per capita CO2 output per country with focus on countries with highest and lowest output suggests that there may be some correlation between GDP and CO2 output per capita. Following this thought I found a data set to test this hypothesis: co2-emissions-and-gdp-per-capita.
Complete data set, after removing missing values contains data of 118 countries. I looked on countries with highest and lowest CO2 output per capita and compared these data with GDP per capita of these countries. **Figure 7**. shows relation between these two parameters.

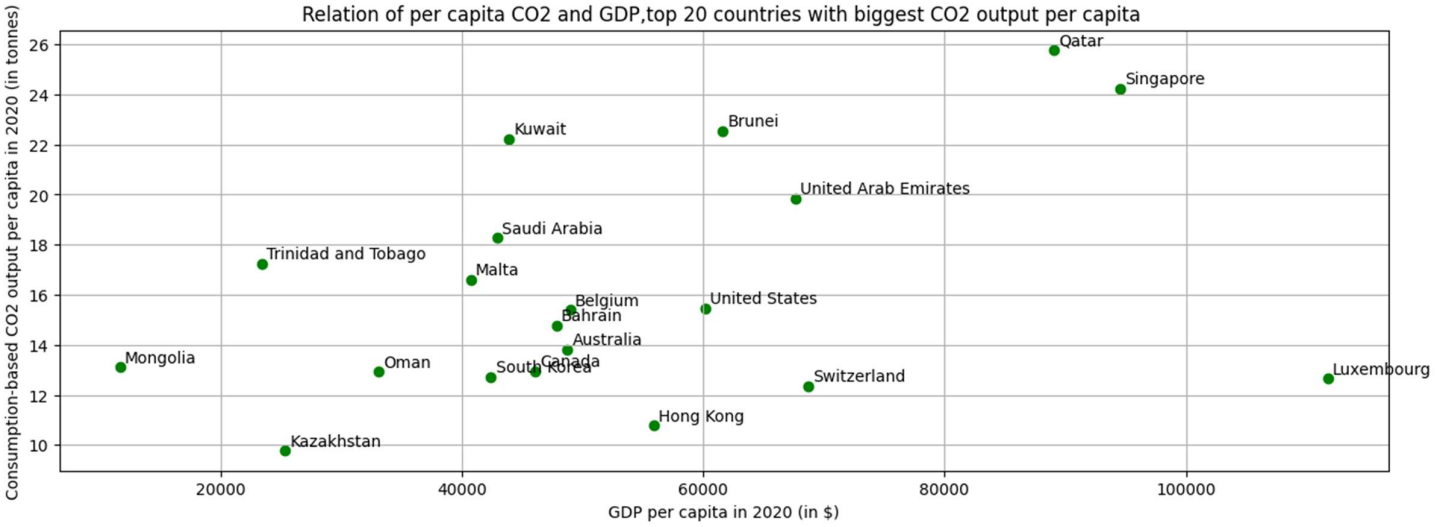


Figure 7. Per capita CO2 output per country in relation with per capita GDP per country. Top 20 countries: highest CO2 per capita per country. Data of year 2020.

We can see that there is obvious correlation, CO2 output per capita increases with GDP, although there are some exceptions, like country Luxembourg. This may suggest GDP is not the only one, strong CO2 output driver.

Figure 8. shows the same comparison for countries with lowest CO2 output per capita in 2020. It looks to be a similar correlation as it was in case of the previously analyzed data.

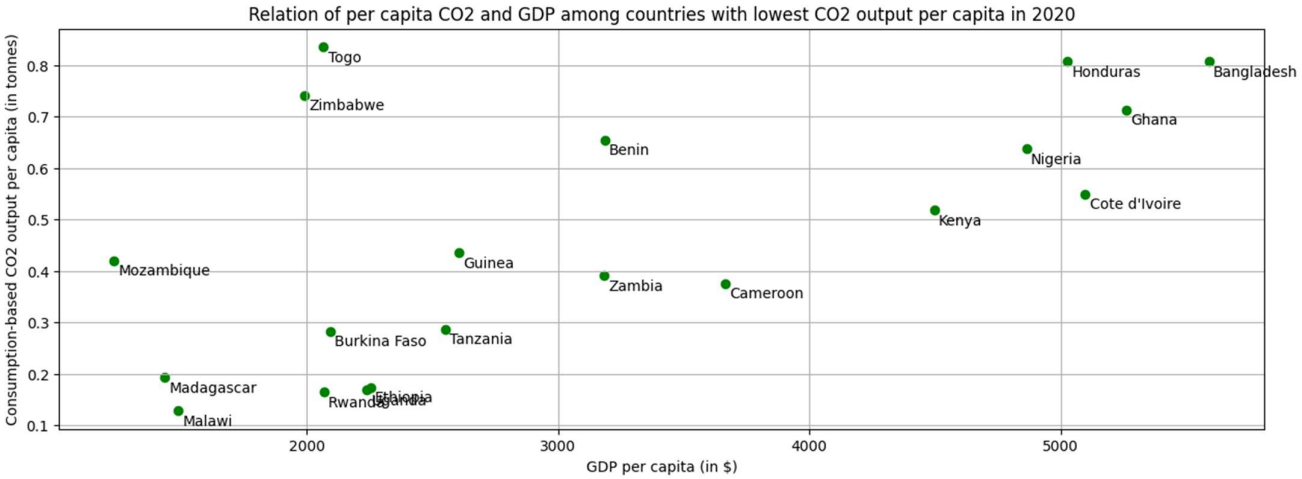


Figure 8. Per capita CO2 output per country in relation with per capita GDP per country. Bottom 20 countries: lowest CO2 per capita per country. Data of year 2020.

To check correlation between GDP and CO2 more closely I have used Pandas function to calculate Pearsons correlation, which makes it possible to indicate how strong that correlation is. Using it I have retrieved a list of countries with strongest correlation of GDP and production CO2, along with consumption CO2. **Figure 9** shows samples of that data.

Entity		Prod_CO₂	Cons_CO₂
India	GDP	0.99	0.99
Vietnam	GDP	0.99	0.99
Cambodia	GDP	0.92	0.98
China	GDP	0.96	0.98
Paraguay	GDP	0.89	0.98

Entity		Prod_CO₂	Cons_CO₂
United Arab Emirates	Year	-0.74	-0.66
United States	Year	-0.87	-0.66
Canada	GDP	-0.65	-0.66
Denmark	GDP	-0.79	-0.68

Figure 9. Pearsons correlation coefficient of GDP per capita to CO2 output per capita. Sample of calculation on group of 118 countries.

We can see that for some of the countries the correlation is the opposite than for others (GDP declines together with CO2). Therefore I have analyzed some of these outliers with strongest positive and negative correlations and plotted data over the years to see how both, GDP and CO2 per capita, are changing. **Figures 10-13** show the results.

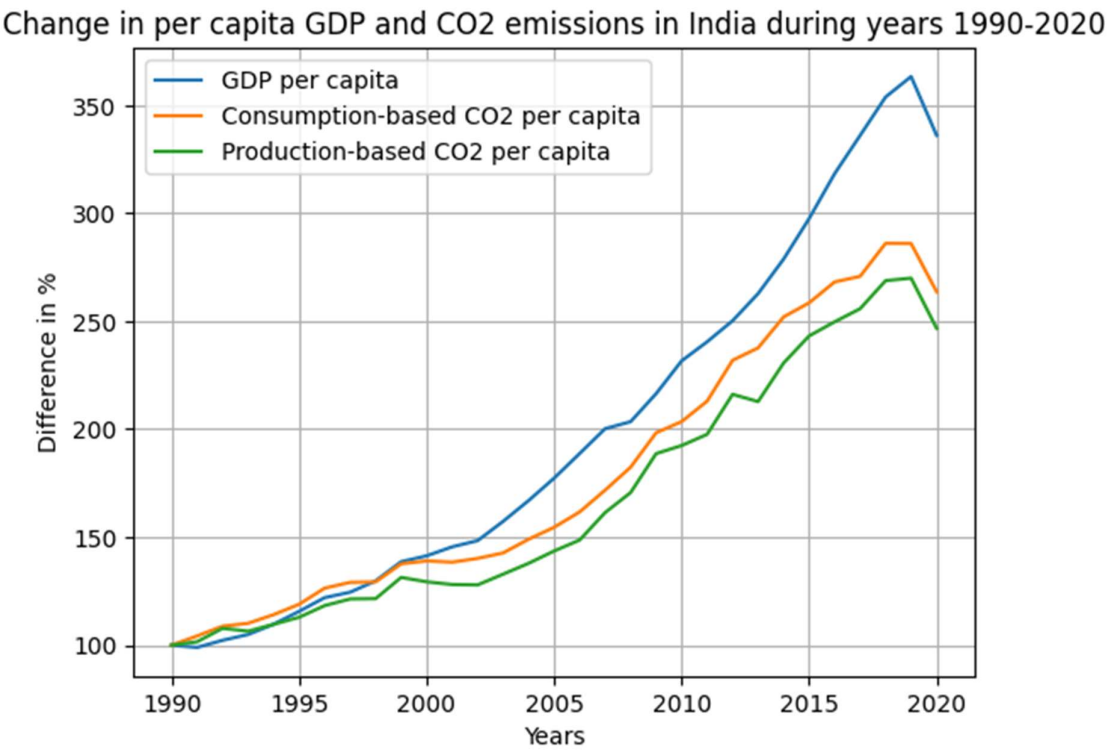


Figure 10. Change in per capita GDP and CO2 output per country. Example of India.

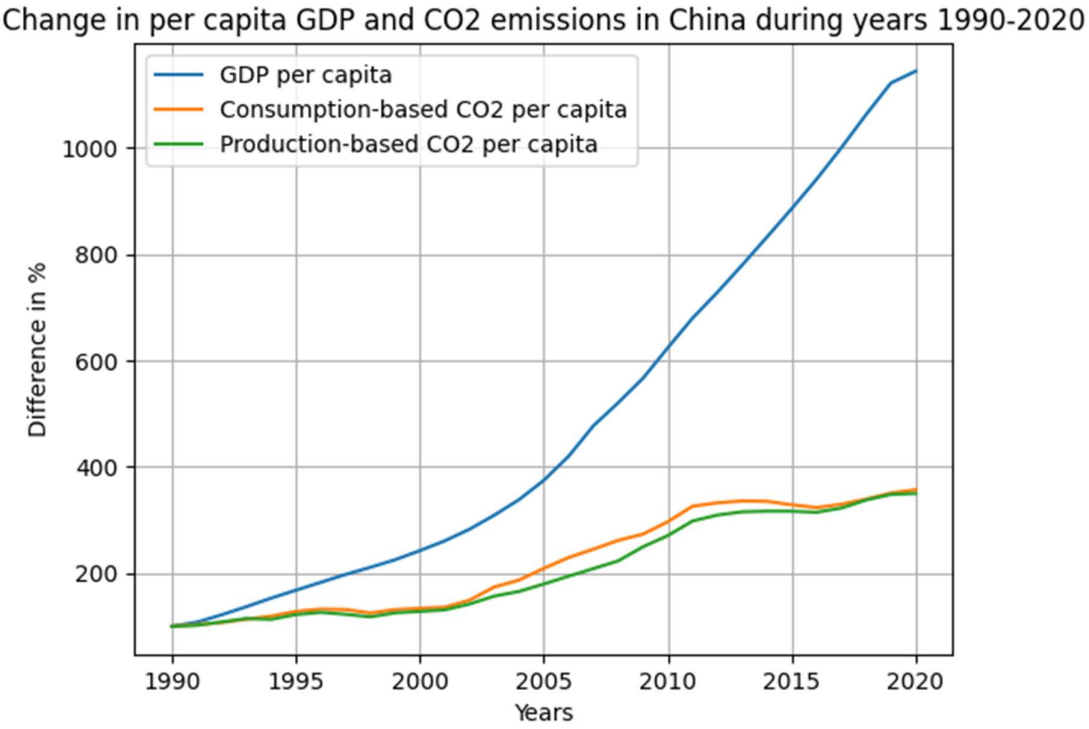


Figure 11. Change in per capita GDP and CO2 output per country. Example of China.

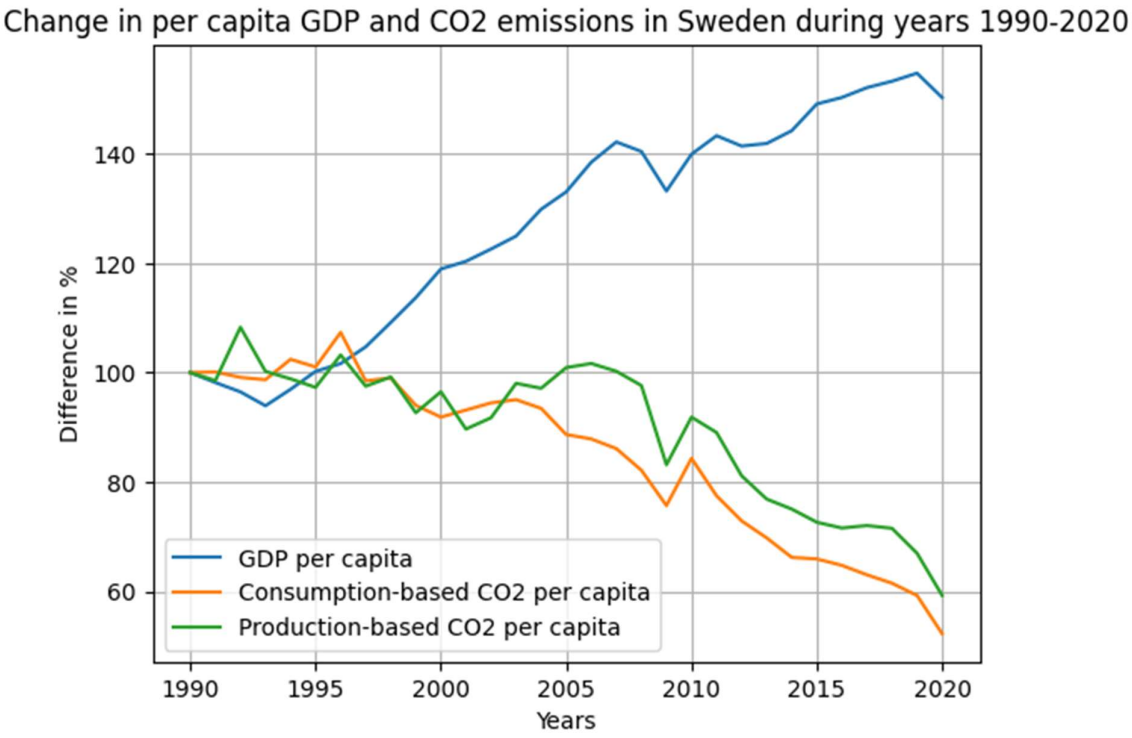


Figure 12. Change in per capita GDP and CO2 output per country. Example of Sweden.

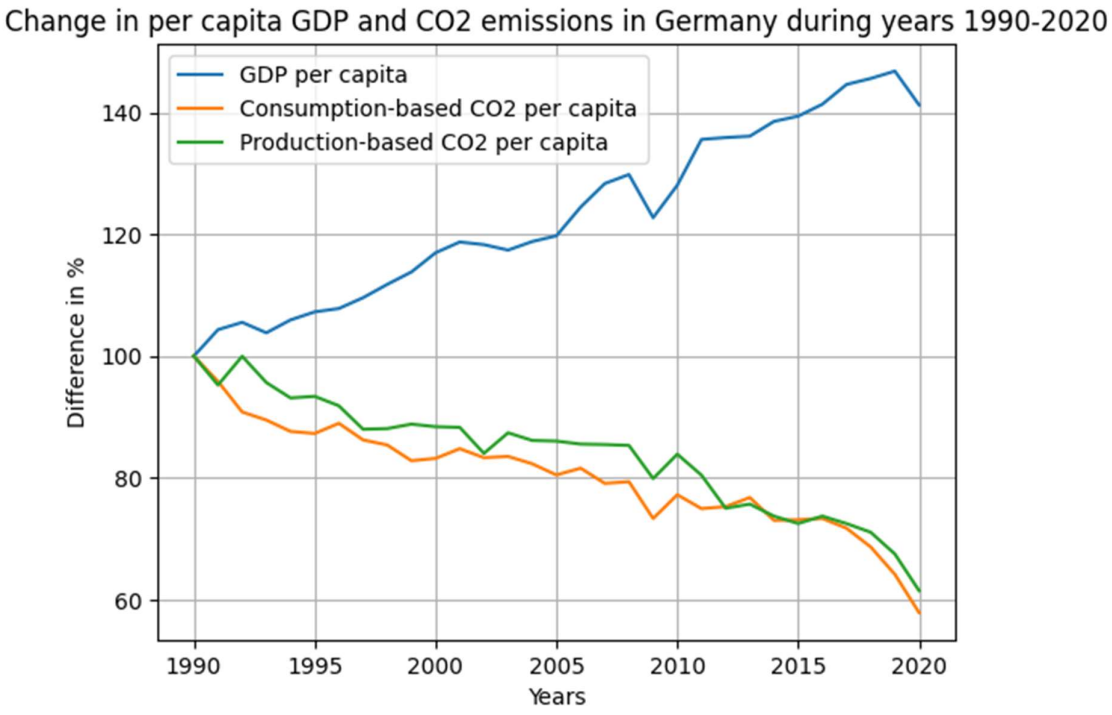


Figure 13. Change in per capita GDP and CO2 output per country. Example of Germany.

GPD per capita has strong influence on CO2 output per capita of a country. It can be positive or negative.

3. **Access to clean fuels for cooking** (codes and results in Jupyter-notebook paragraph 7)

We have already seen that GDP per capita has influence on CO2 emissions per capita. In some of the countries people , which have more money contribute to CO2 emissions increase, in others one opposite is the case. To check this further I decided to analyze data on access to energy (some people may not afford it) and I have found this data:

access-to-clean-fuels-for-cooking-vs-gdp-per-capita.

I have analyzed data of 178 countries in year 2020 and looked at relation between access to clean fuels for cooking and CO2 output per capita. There is a positive relation between GPD and access to clean fuels for cooking but relation to CO2 is not that strong. Therefore I consider access to clean fuels for cooking not one of the biggest drivers of CO2 emissions per capita.

4. **Energy consumption** (codes in Jupyter-notebook paragraph 8)

After analysis on clean fuels for cooking I decided to look for some more general data about energy use and its relation to CO2 output per capita. I have found the data base: 'per-capita-energy-use' and analyzed most recent data of 222 countries.

Results are shown on **figures 14 and 15.**

The more energy consumed, highest CO2 output. But there are outliers, like Norway. So in further chapter I want to look more into details on energy consumption.

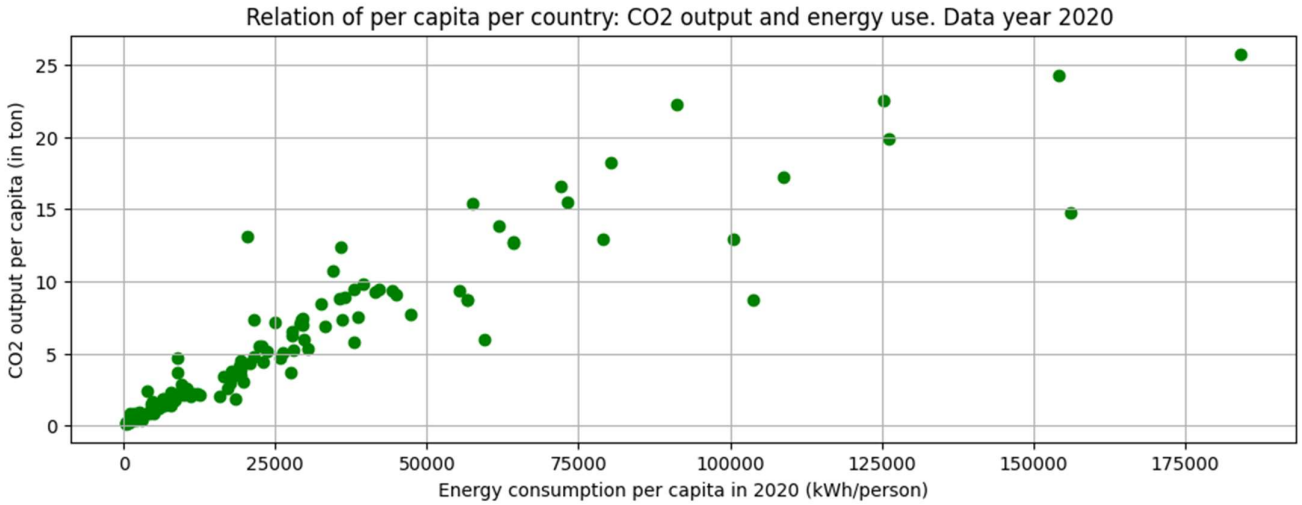


Figure 13. Relation of per capita per country: CO2 output and energy use.

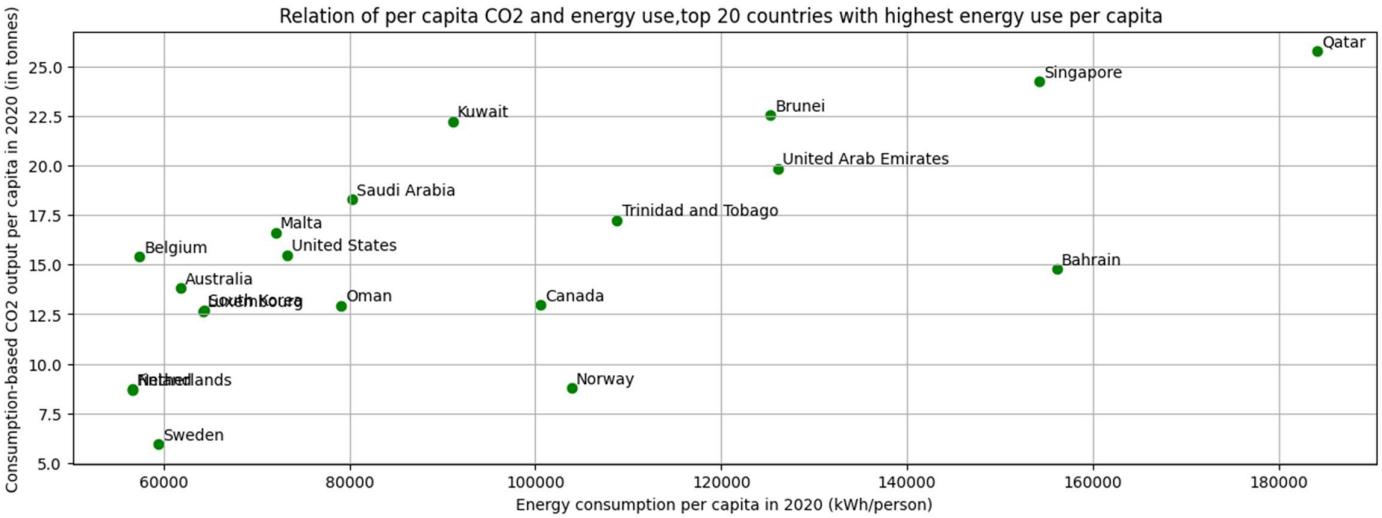


Figure 14. Relation of per capita CO2 and energy use,top 20 countries with highest energy use per capita.

5. **Fossil fuels vs.renewables** (codes in Jupyter-notebook paragraphs 9 and 10)

Looking for more details about energy use I have found two interesting data sets: the use fossil-fuels and use of renewables.

First data set: co-emissions-per-capita-vs-fossil-fuel-consumption-per-capita. After cleaning the data and zooming on 2020 the data of 74 countries left. Results of this analysis are shown on **Figure 15.**

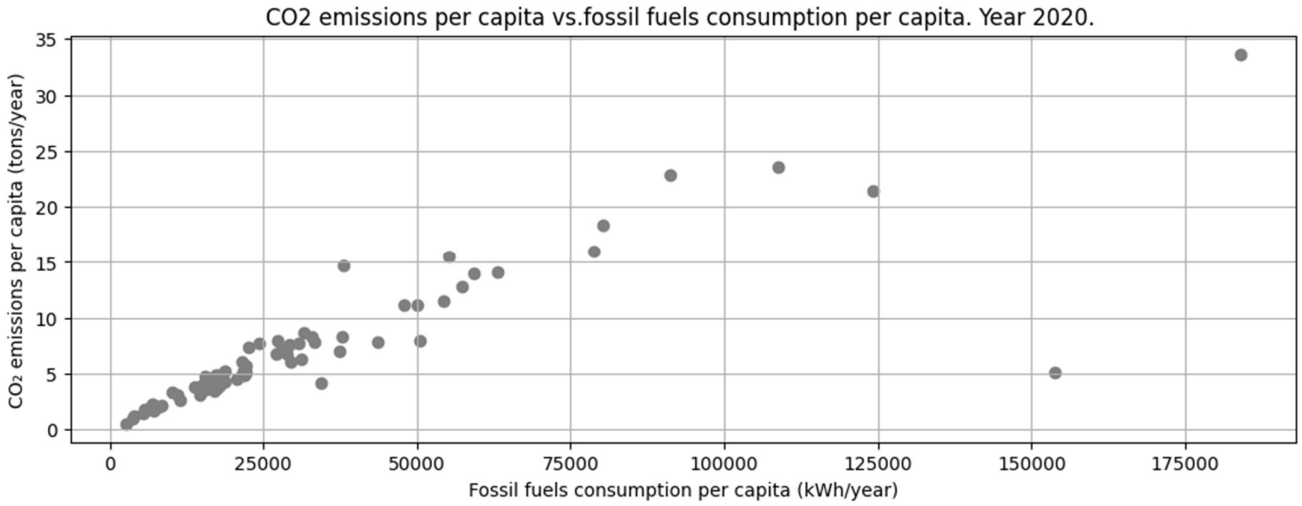


Figure 15. CO2 emissions per capita vs.fossil fuels consumption per capita. Year 2020.

There is definitely a positive correlation . Also Pearsons coefficient of 0.85 confirms that.

Further on I have analyzed CO2 and fossil fuels energy use over the years and plotted details of countries with biggest positive and negative change in use of fossil fuels. **Figures 16 and 17** show 2 examples of these.

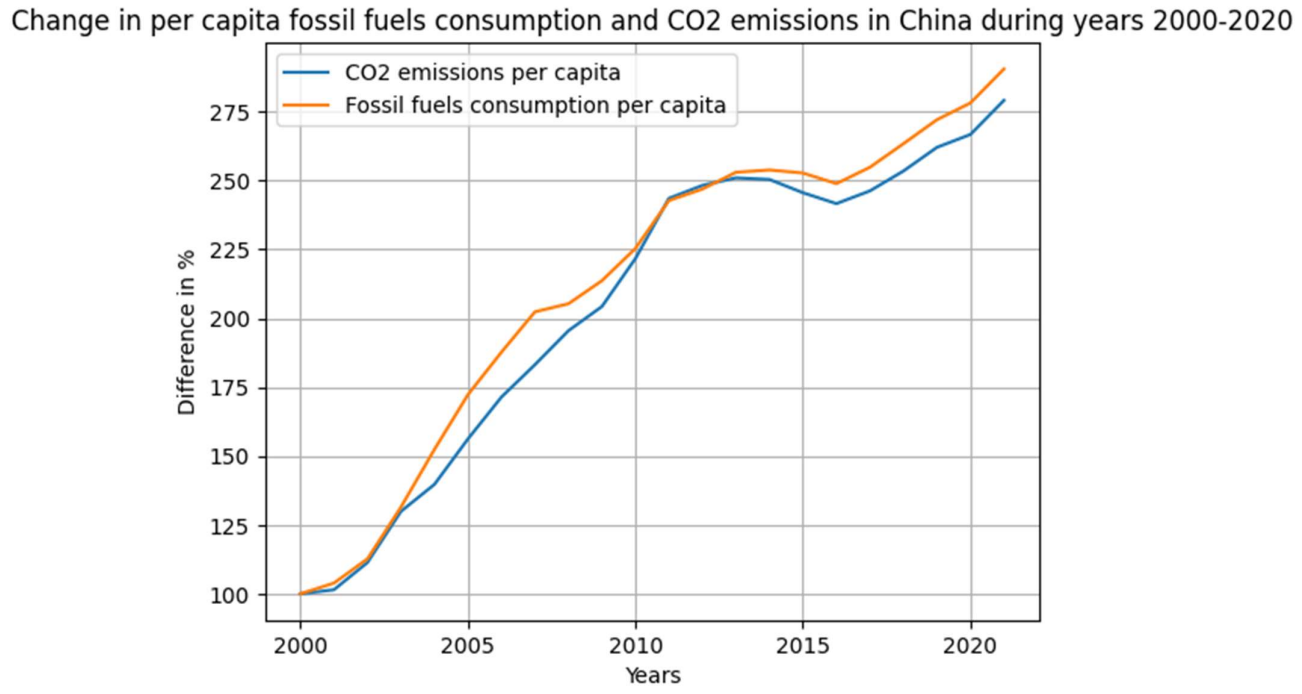


Figure 16. Change in per capita fossil fuels consumption and CO2 emissions in China.

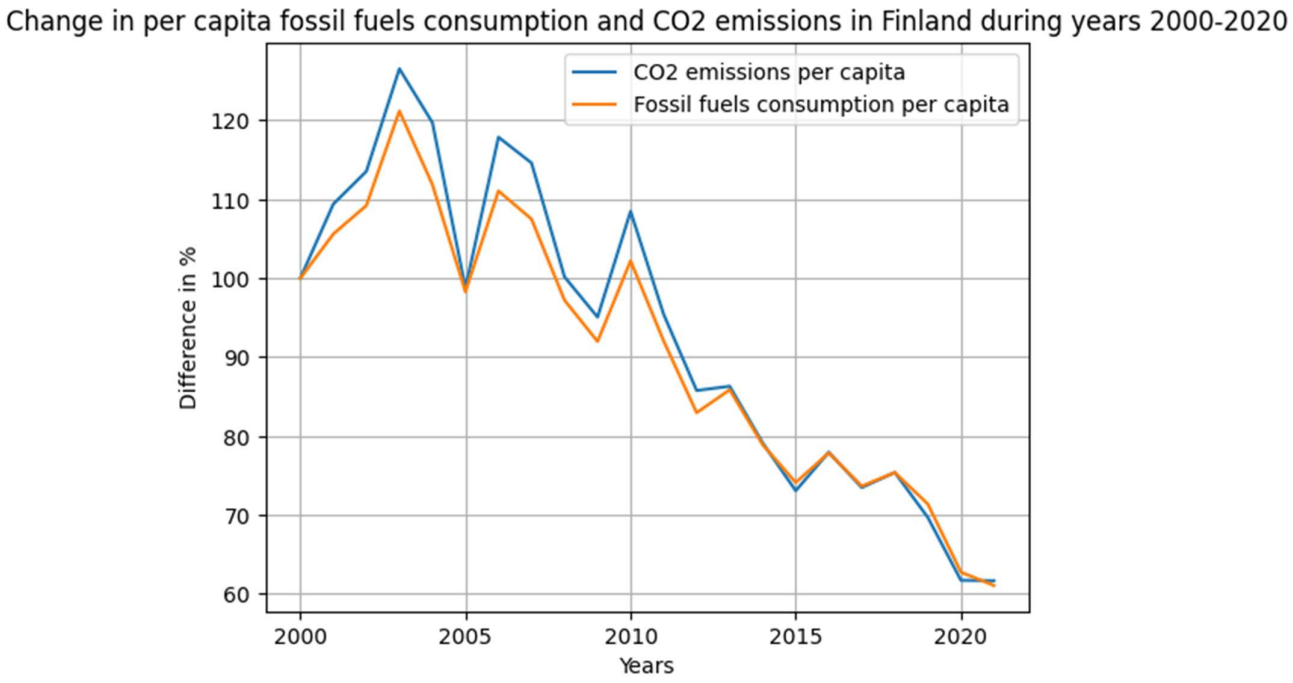


Figure 17. Change in per capita fossil fuels consumption and CO2 emissions in Finland.

Second data set: per-capita-low-carbon-vs-gdp. Available data of 76 countries. First analysis shows moderate correlation of consumption of low-carbon energy and CO2 emissions per capita (Persons coefficient of 0.42). **See figure 18.**

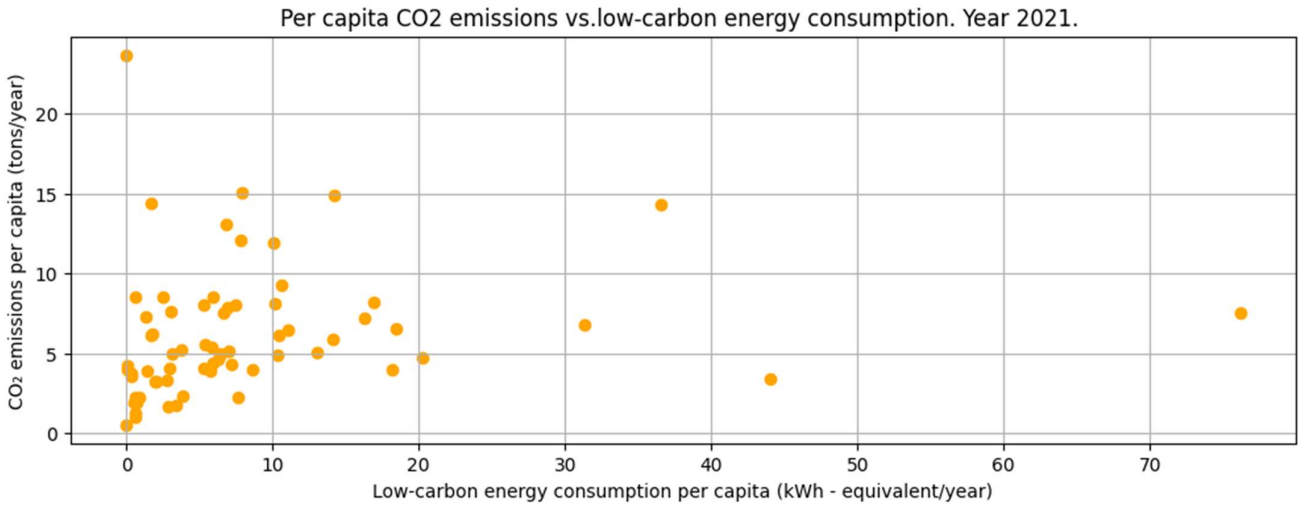


Figure 18. Per capita CO2 emissions vs.low-carbon energy consumption

Looking more into details on country level. I have calculated change in renewables use over the years and looked on the countries with biggest positive and negative changes. Results of this analysis are shown on **Figures 19 and 20.**

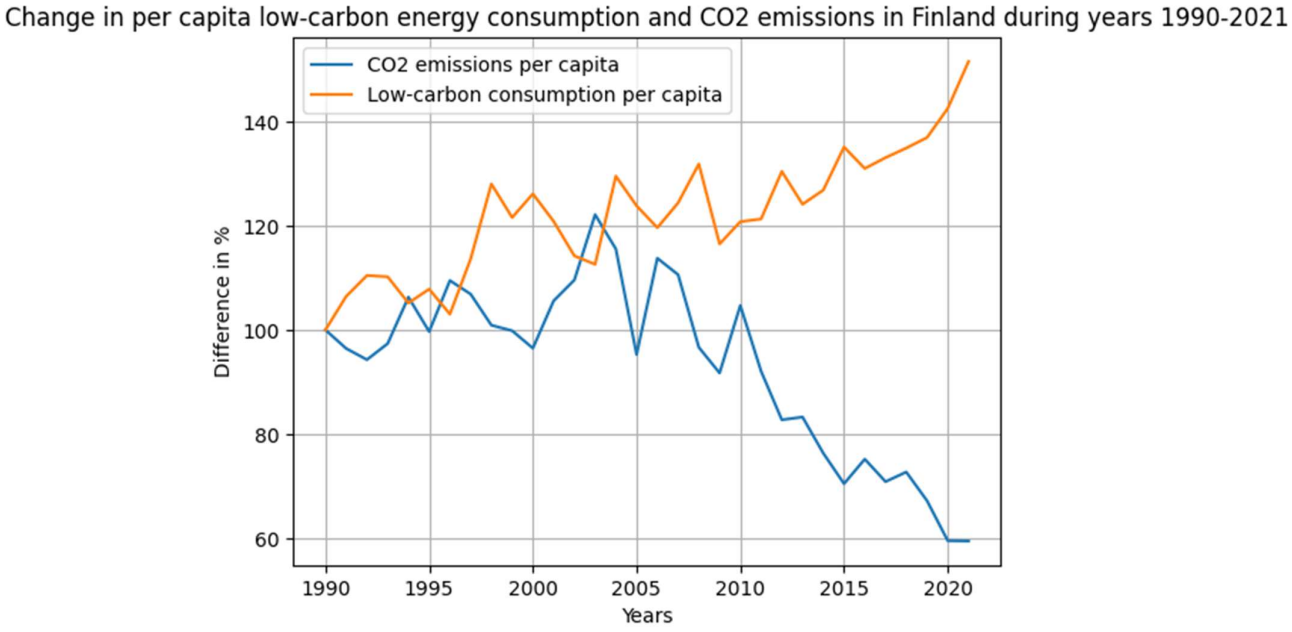


Figure 19. Change in per capita low-carbon energy consumption and CO2 emissions in Finland.

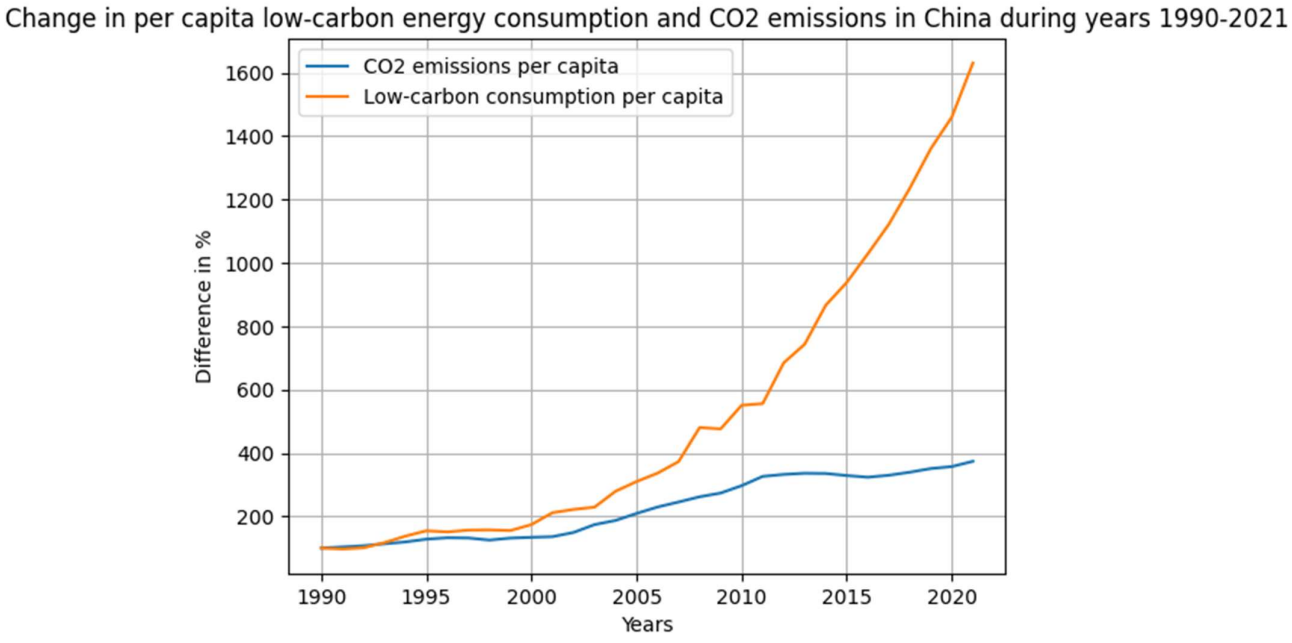


Figure 20. Change in per capita low-carbon energy consumption and CO2 emissions in China.

Example of Finland shows decrease of CO2 output of capita while use of renewable energy consumption per capita increases. But that is not the case for all countries. Example of China shows that this correlation is weaker. This suggest that other factors are of the influence as well and , if we look back on data of fossil fuels we see that use of

these in China is increasing. When we look at other countries and compare both energy sources use we can quickly draw conclusion that combination of those 2 is important driver of CO2 output (per capita).

Final conclusion

Co2 output per capita of country depends on many factors. First of all it depends on country itself and its economics and industry activities, sort of resources which are being used for production and householding, and it's access to resources which have further influence on export etc.). So, it's a complex matter. But there are some visible general trends.

After analysis of available data I can draw conclusion that the strongest drivers of CO2 output per capita are population, GDP per capita and energy use, in specific share of fossil fuels use and low-carbon sources use.

Ad.2 Which countries are making the biggest strides in decreasing CO2 output?

Link to notebook:
https://github.com/MonikaGost/Winc/blob/60dbf749701a3037b1f7ec85c7b94120cfd92162/Final_assignment/Final_assignment_Part2_CO2_changers.ipynb

To answer this questions I have started with analysis of CO2 output per country. I have chosen to look at consumption-based CO2 emissions, so production-based CO2 emissions corrected by emissions embedded in trade as these reflect not only industry activities of a country but also consumption/lifestyle choices of country's citizens.

For the purpose of this analysis I have looked at CO2 output data of 117 countries over the years 1990-2020. Consumption based CO2 data is not available for many countries, especially before year 1990.

Firstly, I have calculated the relative change in CO2 output between years 1990 and 2020 and extracted a list of 10 countries with highest relative change (**figures 21 and 22**).

Year	1990	2020
Entity		
Ukraine	500602270.00	190236340.00
Estonia	31386010.00	12476984.00
Lithuania	43367890.00	20297068.00
Romania	160936340.00	81097070.00
Slovakia	73209360.00	37443450.00
Greece	96367770.00	53118064.00
Bulgaria	61363516.00	36497030.00
Finland	80695360.00	48002290.00
Kazakhstan	289233540.00	185503020.00
Zimbabwe	18000292.00	11613618.00

Figure 21. List of top 10 countries by CO2 relative change between 1990 and 2020

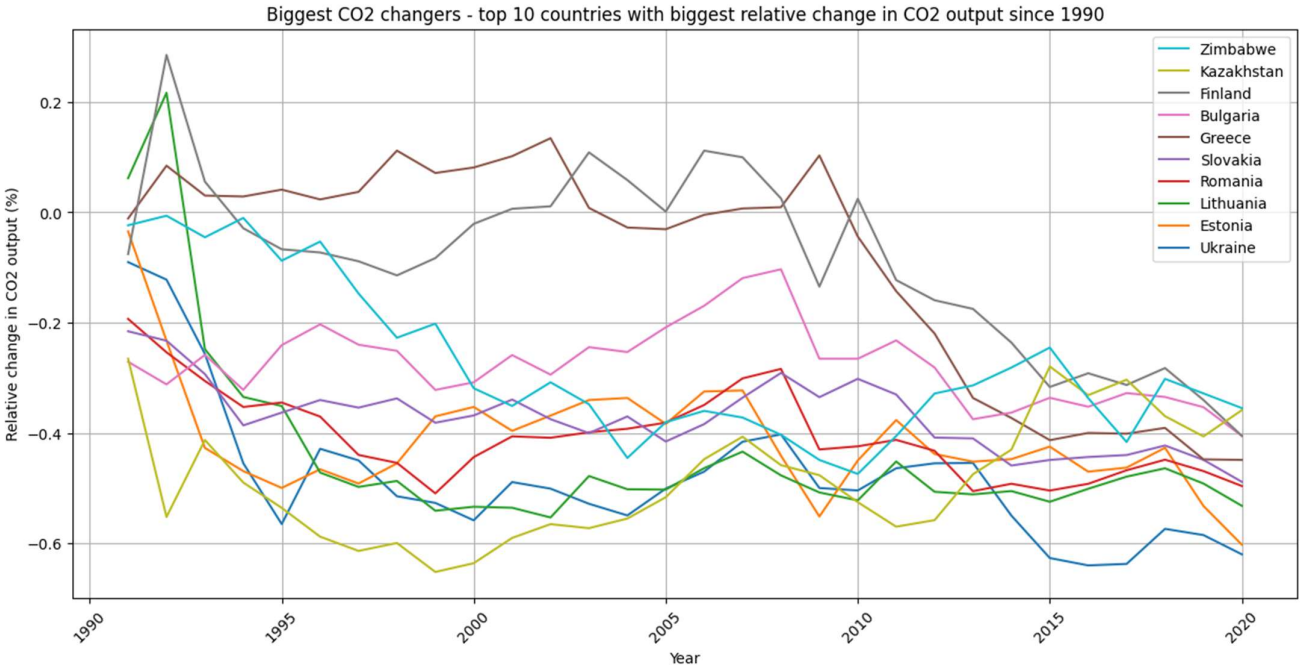


Figure 22. List of top 10 countries by CO2 relative change between 1990 and 2020 over the years.

If we compare these results with population data, we see that some of these countries lowered emissions even though its population was growing. Although for most of the top changers that is not the case. **Figure 23** shows this relation.

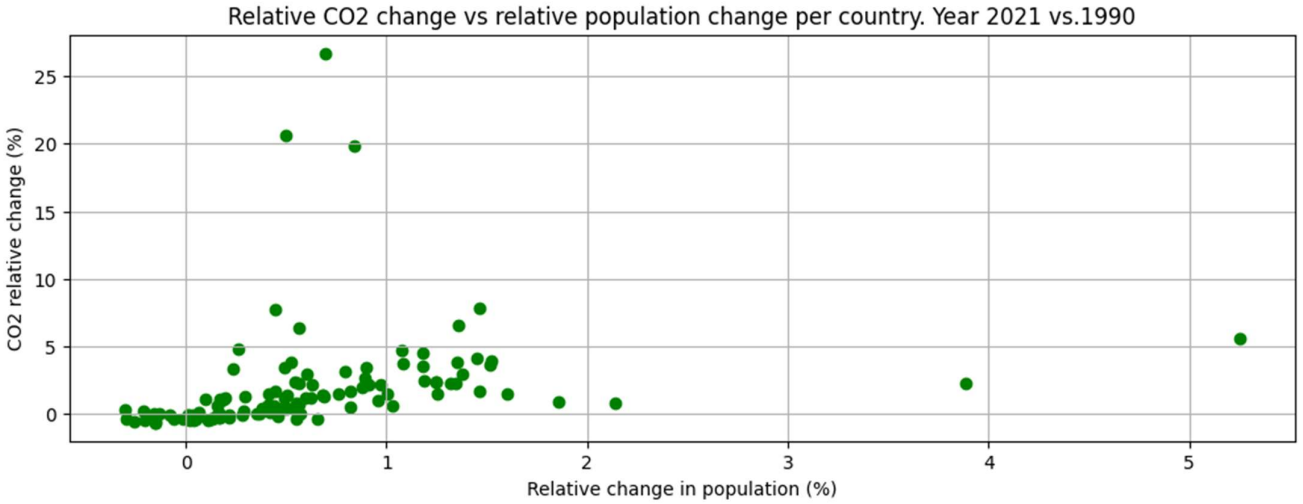


Figure 23. Relative CO2 change vs.relative population change per country.

Also the biggest changer, Ukraine, dropped also in population number. Below figures **24** and **25** show two examples of CO2 drop while shrinking and growing population.

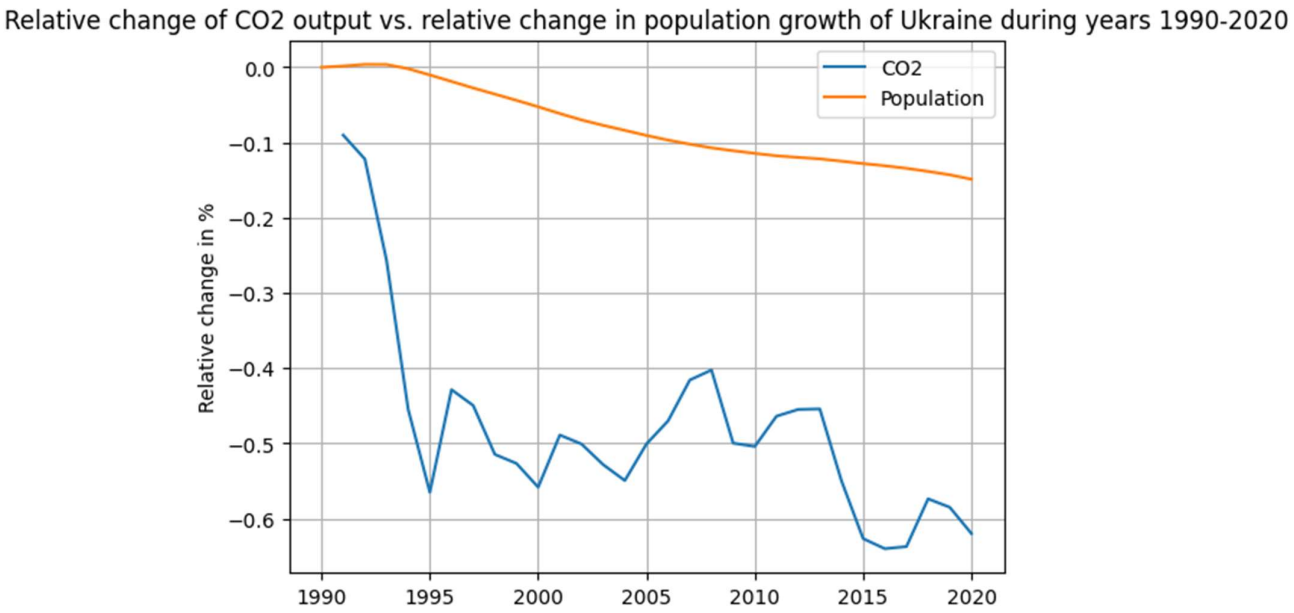


Figure 23. Relative change of CO2 output vs. relative change in population growth of Ukraine.

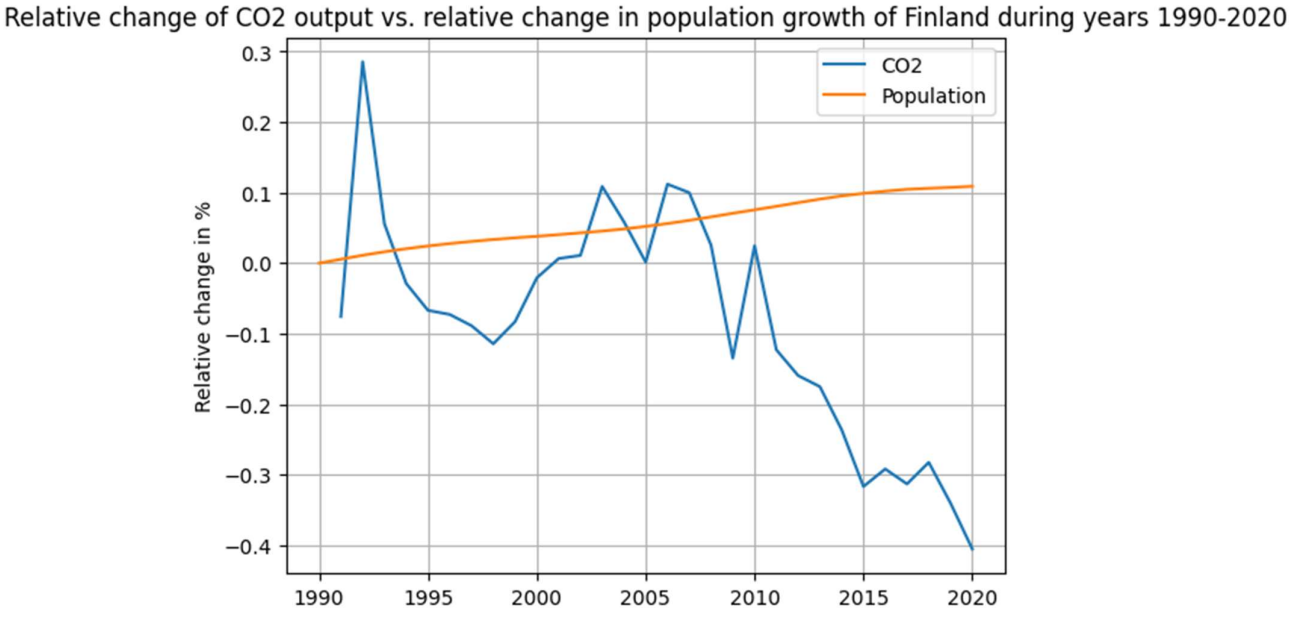
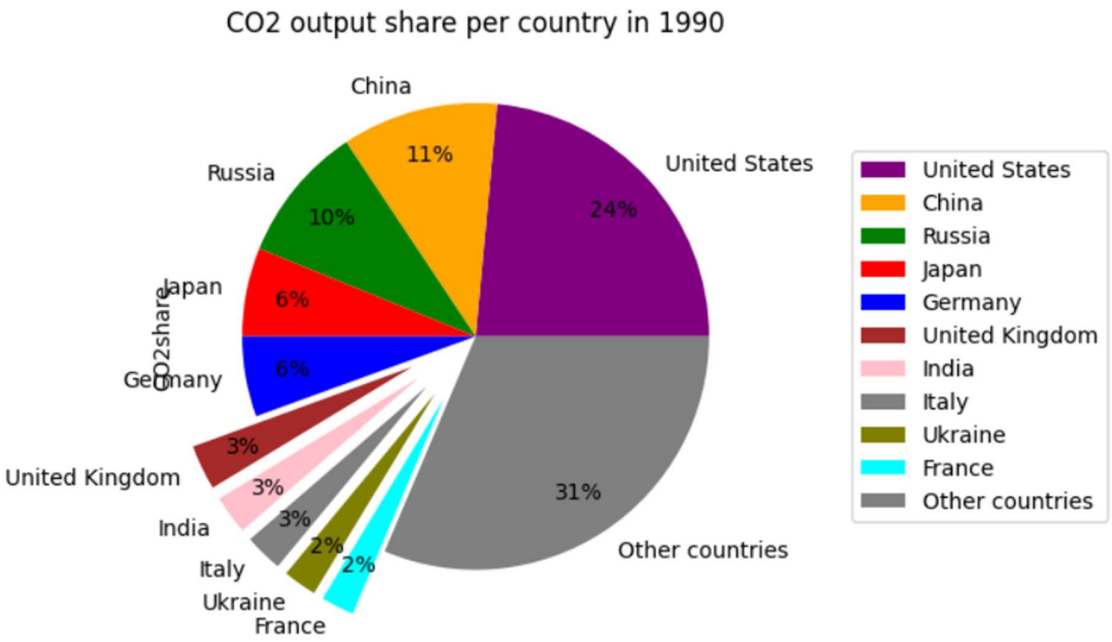
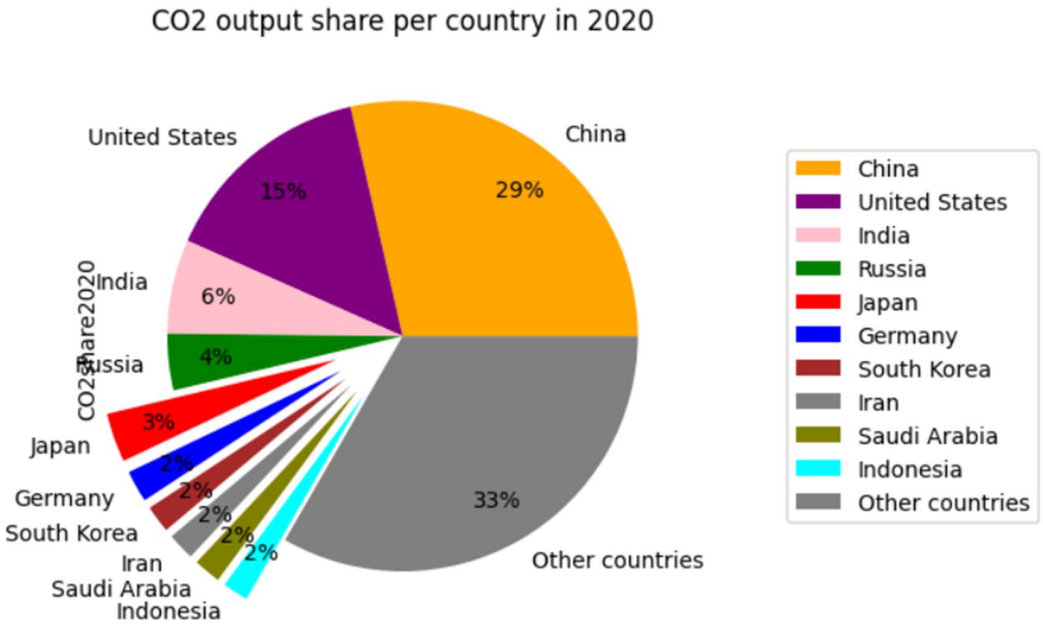


Figure 24. Relative change of CO2 output vs. relative change in population growth in Finland.

I am wondering about biggest CO2 producing countries, as I did not see them among big changers. I make an analysis of that and use the same data here. I look who the biggest emitters are and what is their share in total emissions. Because we are analyzing change I compare start year of this data frame with end year.

Results are shown in below pie-graphs.





What I have found out is, that the biggest polluters are responsible together for ca. 70% of all CO2 emissions. That's why it is interesting to investigate which of these countries have decreases their CO2 output in recent years.

After calculating relative CO2 change for these countries I have plotted the results in **Figure 25.**

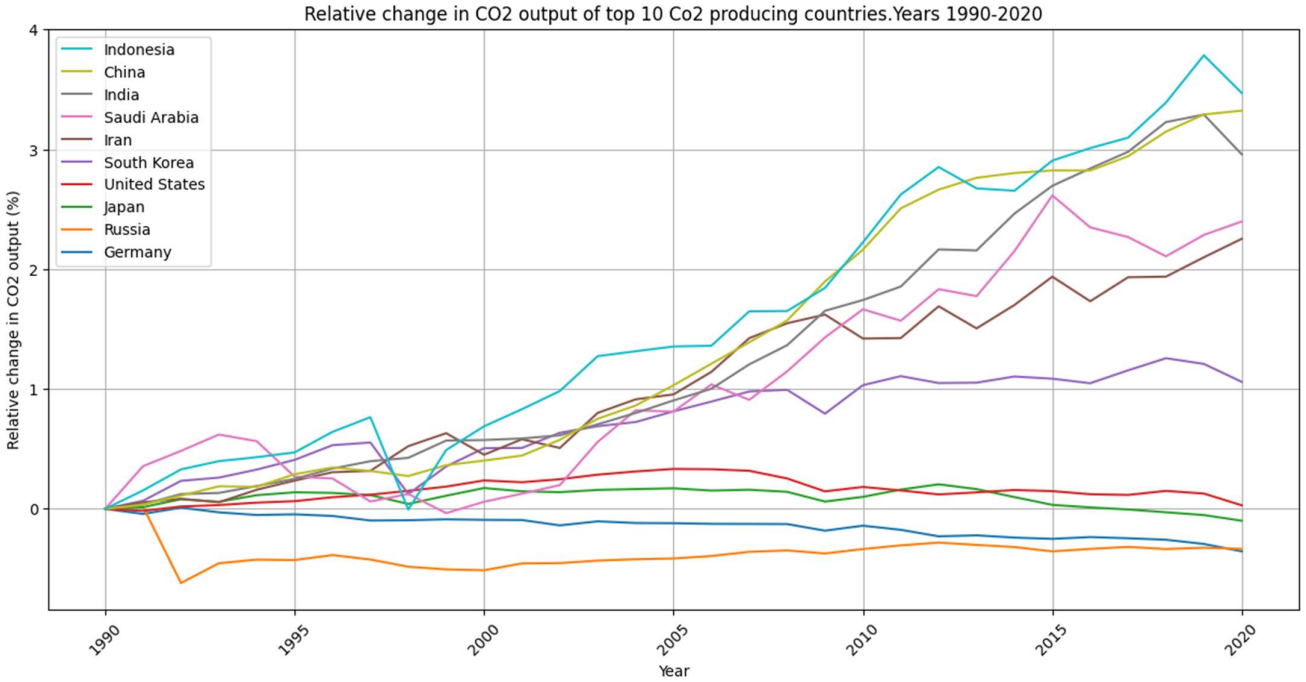
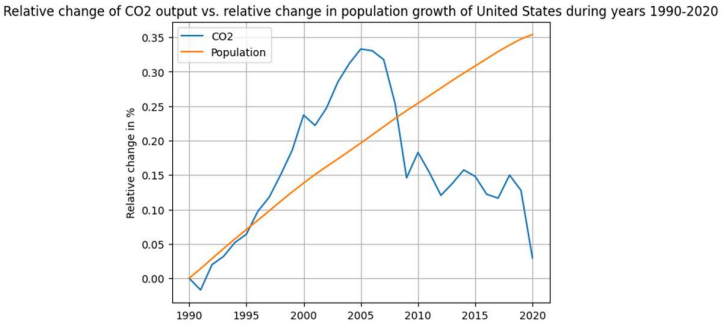
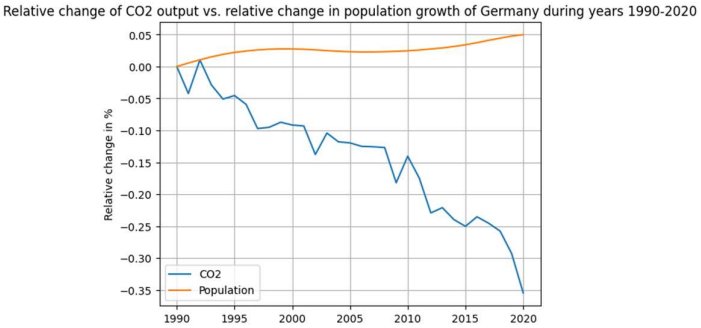
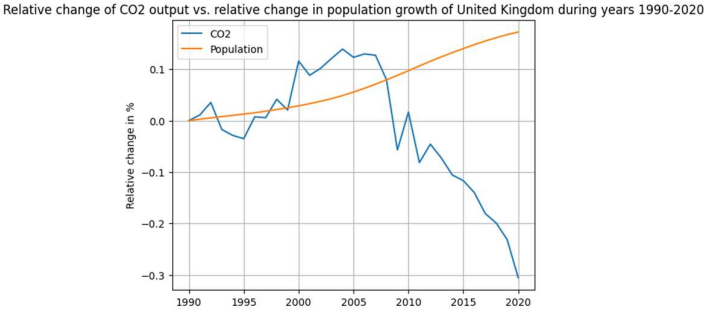


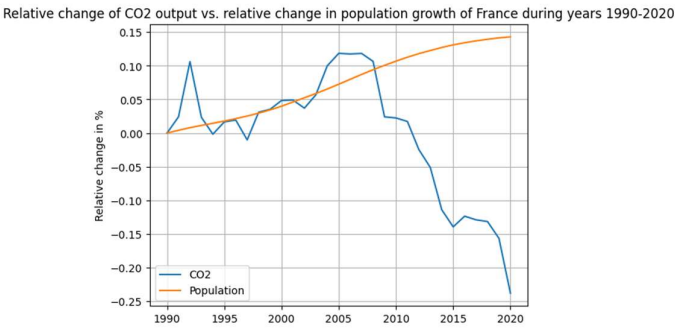
Figure 25. Relative change of CO2 output of top 10 Co2 producers in 2020.

Furthermore I have calculated and plotted the data of population change of these countries. It occurs that several of them, like France, United Kingdom, Germany and United States have dropped emissions aside growing population. That is a very good achievement. Below 4 charts visualize that.









Final conclusion.

In the analysis of biggest CO2 strides I have compared countries with biggest relative change but it turned out very quick that these countries are not te top polluters of our world. From these countries Ukraine has a biggest relative CO2 emissions change but next to that it's population has decreased over the past years. From this group of countries Finland, Kazakhstan and Zimbabwe seem to be achieve CO2 decrease while their populations are growing.

Second part of the analysis concerned the biggest 10 polluters, which are together responsible for over 70% of global CO2 emissions. So their CO2 decreases really make a difference. From these countries Germany, Russia and Japan have booked biggest relative decreases while countries like France, United Kingdom, United States and again Germany are booking them too despite growing populations.

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

Link to notebook:

https://github.com/MonikaGost/Winc/blob/60dbf749701a3037b1f7ec85c7b94120cfd92162/Final_assignment/Final_assignment_part3_future_price.ipynb

To answer this question I have analyzed data set: levelized-cost-of-energy-by-technology which contains the average cost per unit of energy generated across the lifetime of a new power plant.

There was not much data available on this matter and in the given data set the data prices for most of the technologies are from last 10-20 years.

The only variable to compare prices with is year, so we base future price on the available price data of the past years only. No other circumstances are being taken into consideration in this analysis.

For each technology I have looked on the coefficient of correlation to estimate how reliable our prognosis is.

Secondly linear regression model was fitted to predict the future price – in year 2030.

I have gathered calculated future prices in below table:

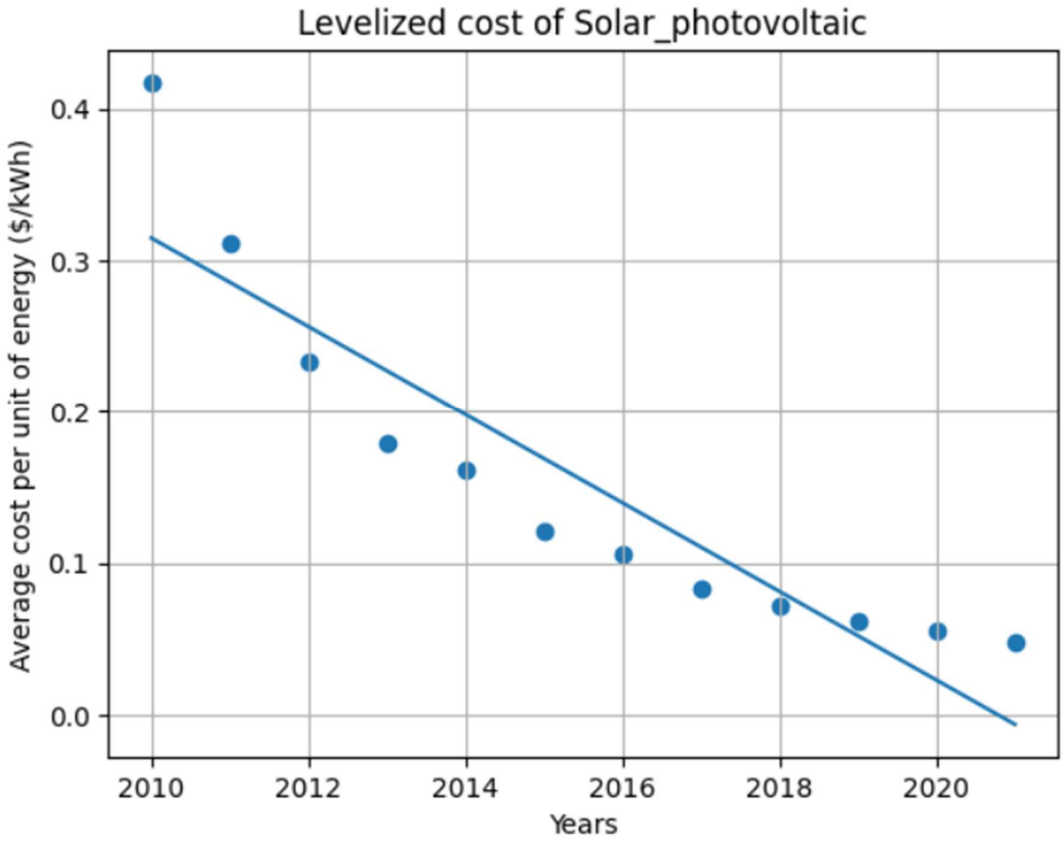
*** Predicted prices of non-fossil energy sources in 2030 ***

Energy_source	2030_price
Offshore_wind	0.08
Geothermal	0.07
Bioenergy	0.06
Hydropower	0.05
Onshore_wind	-0.04
Concentrated_solar_power	-0.09
Solar_photovoltaic	-0.27

The price if solar photovoltaic does not look realistic but it will be with big probability the best price in the future.

Below figure shows historical pricing data of Solar photovoltaic and fitted linear regression line.

```
*** The correlation between years and Solar_photovoltaic is:
-0.9148528097433445
```




```
*** Predicted price of Solar_photovoltaic in 2030 is: -0.26877645473309997 $/kWh ***
```

End of report.