# Final assignment - CO2 emissions

## Introduction

This report describes the results of the three questions posed for the final assignment for the Data Analysis with Python course by Winc Academy.

To answer the questions datasets were taken from [Our World in Data website](https://ourworldindata.org/).

The three questions to be answered are:

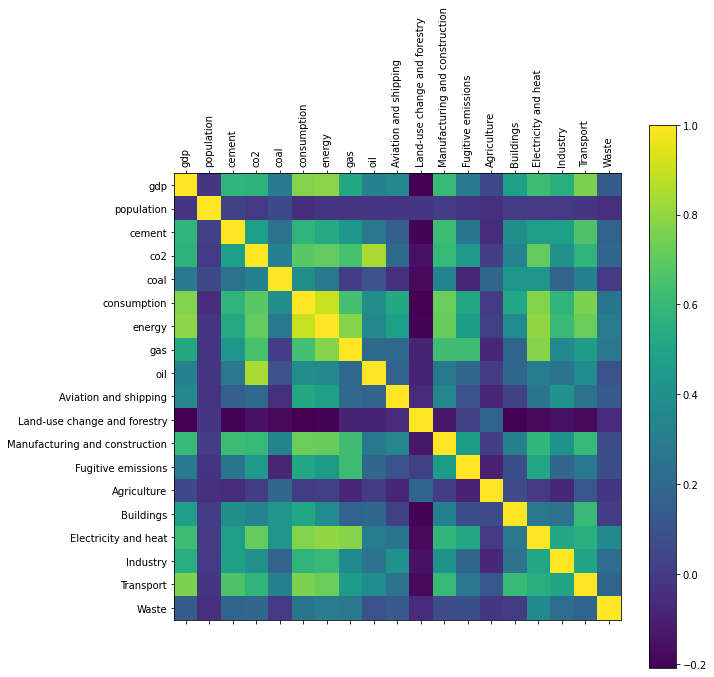
1. What is the biggest predictor of a large CO2 output per capita of a country? *Page 2*
2. Which countries are making the biggest strides in decreasing CO2 output? *Page 3*
3. Which non-fossil fuel energy technology will have the best price in the future? *Page 4*

Each question has a corresponding notebook for the code with the following names: Final Assignment Q1, Final Assignment Q2 and Final Assignment Q3, respectively.

## 1: Biggest predictor of CO2 output

One of the datasets I found showed CO2 emissions mostly from fossil fuels and greenhouse gases. Another data set showed released CO2 emission equivalents produced in certain sectors (i.e. Aviation and shipping, Manufacturing, etc). To determine the biggest predictor of CO2 output, the correlation coefficients were calculated between the different factors and sectors. Only data equivalent to CO2 emissions ‘per capita’ were taken into account. The correlation coefficients were visualized in a 2D matrix as color-coded image (Figure 1).

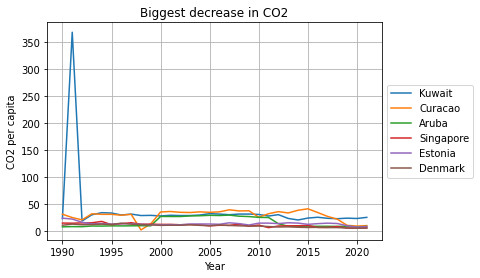
Based on the correlations between these different factors and sectors, it becomes apparent that there is a high correlation coefficient between CO2 emissions and ‘energy’, ‘consumption’ and ‘electricity and heat’ (each about 0.9). Therefore the biggest predictor for high CO2 emissions per country are the energy production and consumption per capita, related to electricity and heat. GDP and CO2 emission also has a high correlation coefficient (0.7), countries with a higher GDP per capita also tends to have higher CO2 emissions.

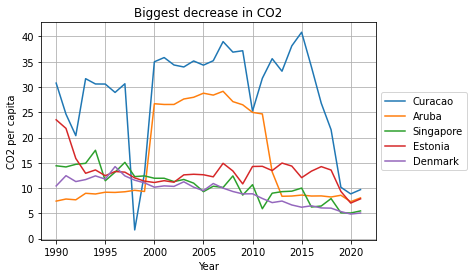
Figure 1: Correlation coefficients of factors and greenhouse gas sectors producing CO2 output. The correlation coefficients were calculated on CO2 outpput per capita. Numbers closer to 1 indicate a possible positive correlation, while numbers close to 0 indicate that there is no correlation.

## 2: Biggest strides in decreasing CO2 output

To find the relatively biggest decrease in CO2 output, I decided to look at the relative change of 2021 (most recent data), compared to the maximum output since 1990. Before this date many countries were lacking (trustworthy) data. To take into account that countries can have growing and shrinking populations too, I used the CO2 output per capita (or per person) of a country.

The relatively lowest outputs for the five countries Kuwait, Curacao, Aruba, Singapore, Estonia and Denmark varied from 7 – 33% of their maximum output (Figure 2). However, from the plot it could be seen that Kuwait showed a sharp high peak around 1991, most likely due to an outlier, and giving a false result when calculating the relative decreasing trend. Therefore the fifth country making biggest strides in decreasing CO2 output per capita would be Denmark (35% of maximum output). Figure 3 shows the developments of the CO2 emission for the top five countries that showed biggest strides in decreasing their CO2 output.

Figure 2: CO2 output per capita over the past *3*0 years are plotted for the top five countries that have highest relative decrease in CO2 output. Initially Kuwait showed up, most likely due to an outlier around 1990.

Figure 3: CO2 output per capita over the past 30 years for the top five countries that showed biggest strides in decreasing their CO2 output.

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

For the question which non-fossil fuel energy will have the best price in the future I used a dataset containing levelized cost of energy by technology, assuming that the price will be linearly related to the costs.

I have calculated and plotted the simple linear regression lines to show the trends and get an indication on future prices. The r squared gives an indication on the fit of the regression lines, which in most cases is very poor (the closer to 1 the better). Possibly higher degree polynomial regression would result in better fits with the data, but due to the very small size of the data set, I decided to keep the simple linear regression lines for an easy comparison on predicted prices. A summary of the results is shown in Table 1. The table also includes the predicted future prices for 2022.

Based on the results as shown in the table the price for photovoltaic solar power is expected to have the lowest price, followed by onshore wind energy.

Side notes: The regression lines will result in negative values over time for bioenergy, wind and solar power. Onshore wind has the steepest slope, indicating most rapid decline in costs. For photovoltaic solar power the plot already shows a less steep slope since 2018, indicating a slower decline.

Table 1: Results of the linear regression lines for the different non-fossil fuel energies. Also included are the predicted future prices for 2022.

| **Energy** | **Regression line**  **(Costs = a \* Year + b; a, b)** | **r2** | **Costs in 2022**  **(US$/kwh)** |
| --- | --- | --- | --- |
| Bioenergy | -4.63e-04, 1.00 | 0.0347 | 0.067 |
| Geothermal | 6.916e-05, -7.353e-02 | 0.0006 | 0.066 |
| Offshore wind | -3.74e-03, 7.672 | 0.3446 | 0.107 |
| Solar photovoltaic | -2.917e-02, 5.894e+01 | 0.837 | -0.035 |
| Concentrated solar power | -2.177e-02, 4.412e+01 | 0.8638 | 0.089 |
| Hydropower | 8.219e-04, -1.614 | 0.3721 | 0.048 |
| Onshore wind | -6.612e-03, 1.338e+01 | 0.9355 | 0.016 |