**Data Analysist Report on CO2 emission**

*Title Page: CO2 predictor, decrease and renewable energy*

Prepared by: *Joëlle Kappert*

Organization: *Winc Academy*

Date*: February 28, 2025*

Table of Contents

[*1. Executive Summary* 1](#_Toc191645840)

[*2.* *Introduction* 1](#_Toc191645841)

[*3. Methodology* *and* *definition* *of* *choices* 2](#_Toc191645842)

[*4. Data* *Analysis* 3](#_Toc191645843)

[*5. Result* 3](#_Toc191645844)

[5.1 What is the biggest predictor of a large CO2 output per capita of a country? 3](#_Toc191645845)

[5.2 Which countries are making the biggest strides in decreasing CO2 output? 4](#_Toc191645846)

[5.3 Which non-fossil fuel energy technology will have the best price in the future? 5](#_Toc191645847)

[*6.* *Conclusion* 5](#_Toc191645848)

[*7. Appendices* 5](#_Toc191645849)

[*8.* *References* 5](#_Toc191645850)

## 1. Executive Summary

This report provides an in-depth analysis of CO2 emissions from data analyses and research organisation Our World of Data. The objective is to discover which factors have an impact on CO2 emissions output, such as GDP per capita, transport, fossil fuel consumption and population among others. And what actions have been taken to reduce the emission output.

Investments are increasing in renewable energy technologies all over the world, due to the Paris Accord.

## 2. Introduction

CO2 emission is based on greenhouse gas. This report examines data from the past 5 years (2019-2023) to see the movement of the emission output, other influential impact, and the correlation between these elements. The final goal is to visualize the outcomes of the data analysis in order to answer the questions about CO2 emission output, steps taken to reduce emission output and the future of renewable energy technologies.

## 3. Methodology and definition of choices

Analytical tools such as Python and Pandas were used to analyse and interpret the different, large datasets in Google Colab Notebook. MatplotLib was used to visualize the outcomes into graphs (plots).

Multiple large datasets were selected to analyse and find the results requested. The number of cars per capita was outdated (till 2017), this was replaced by the transport dataset. Car sales of electric cars was used for the reduction of emission output analysis because this correlates between having an active plan for reduction and renewable energy.

The first step in analysing datasets is to clean/organize them. This consists of dropping unnecessary columns, for example country codes and countries that are not actually countries, like Africa, Asia, etc. Renaming columns and rounding numbers to make the datasets easier to read and selecting the past 5 years (2019-2023).

After cleaning, checking for missing data (NaN, null or 0.0) was executed. If the missing data was less than 6%, the decision was made to do nothing with it. In some cases it was necessary to drop the missing data in order to combine (concatenate) the results for the graphs. This can be seen from the coding used for concatenating (*pd.concat()*). To get the results, three methods were used to make sure the outcomes were similar. These methods are: groupby in combination with aggregation, idxmax or idxmin and nlargest or nsmallest.

In some cases, additional analyses/calculations had to be made. For example, the year 2023 was missing in the dataset of “GDP per capita”. For the dataset “CO2 emission Transport” the years 2022 and 2023 were missing. These were added by calculating the mean (average) and finally used for the graph ’ Qatar’. Another example is with the dataset on “Dietary”. The column names here were exceptionally long and therefore a different code is used (see Notebook CO2 *Output).*

For the datasets *“Fossil fuels per type” and “Renewable energy production”* a new column ‘Total’ is added. This made it easier to analyse the top or last 5. The same was done for ‘*Electric car sales*’, the added column calculated the relative of total sales (%).

Concatenating different datasets before beginning the analysis, can cause more missing data. Therefore, I used pre-combined dataset between ‘CO2 emission’, ‘GDP’ and ‘Population’ and dropped the ‘GDP’ column since it was not necessary for this analysis. In order to gather the data for the chart, I also dropped the missing data. Struggling with creating the chart because of the use of relative numbers instead of regular numbers, has resulted in displaying only the CO2 decrease without the population.

A map of countries with active CO2 emission reduction plans is included in the Notebook as a reference which I downloaded from Our World in Data/United Nations (see Notebook *Active decrease* CO2*).*

For the last datasets 2 different charts were made. The linear regression chart is displayed in this report. The investment dataset revealed only the world as country, therefore the costs of renewable energy technologies also shows the world (see Notebook *Renew.Tech*).

## 4. Data Analysis

Factors analysed from 2019-2023:

**CO2** **emission per capita**: Qatar has the largest carbon dioxide output of an average 20%. growth.

**GDP per capita:** The GDP has reduced over the years, Qatar still on top.

**Fossil fuels:** China has the largest fossil fuel consumption per fuel type, Qatar has the largest per capita.

**Population:** The largest world population is between China and India.

**Dietary:** The devidance per category and the outcome differs. For example China has the largest vegetables composition, where the USA has the largest vegetable oils and sugar & sweeteners, Bangladesh the largest rice, etc.

**Transport:** China and the USA share the largest CO2 emission output caused by transport.

**Electric car sales:** From the total cars sold, Norway sold 93% electric cars, followed by China with 38%.

**Renewable energy technology:** China and the United States have the largest total production of renewable energy. Iceland has the largest renewable energy per capita. Investments are on-going, prices drop. On-shore wind energy has lowest price.

## 5. Result

### 5.1 What is the biggest predictor of a large CO2 output per capita of a country?

Qatar has the largest CO2 emission output, GDP and fossil fuel consumption per capita. China and USA have the largest fossil fuel consumption per type, CO2 emissions from transport, dietary composition, and population. This analysis shows that the population has impact on data per capita. For example, the larger the population, the lower the GDP, the larger the dietary composition and therefore more transport is required (incl. private cars, public, industrial and aviation).

A graph of a graph showing the number of co2 output

AI-generated content may be incorrect.

Chart: *Largest* CO2 *output*: **Qatar**

### 5.2 Which countries are making the biggest strides in decreasing CO2 output?

The Paris Accord was signed by 195 countries/nations in 2016, meaning all of these countries are actively reducing their carbon dioxide emissions and there is no more difference between developed and developing countries. (source: *United Nations* (UN))

For this topic, decrease in CO2 output, analyses are made for 2023 only. Starting with the relative CO2 emission versus population, the top 5 countries are from Europe (see chart below). Bulgaria has the highest relative reduction of -23% and a population reduction of -3%.

The lowest consumption of fossil fuels (kWh) per capita is by Bangladesh, whereas the fossil fuels by fuel type (total) is by Latvia.

Analysing the car sales for electric/hybrid vehicles, Norway stands out with 93% of the total cars sold that were electric or hybrid, in total 110.000 cars. The highest total of electric cars sold is 8.100.000 (8.1 million) by China which is 38% of the total amount of cars sold in 2023.

Analysis on renewable energy technologies reveals that China has the largest production of all renewable energy sources. The USA follows on wind- and solar energy. Iceland is the largest with renewable energy per capita, followed by Norway.

The overall outcome per major continent is that China , USA and Scandinavian countries exceed in actively reducing their CO2 emissions output by investing, producing and distributing renewable energy.

A graph of the co2 emissions

AI-generated content may be incorrect.

Chart: *Largest decrease* CO2 *in 2023*

### 5.3 Which non-fossil fuel energy technology will have the best price in the future?

The graph below shows that onshore wind technology has the lowest cost and therefore the best price, followed by solar photocvoltaic energy.

A graph of energy costs

AI-generated content may be incorrect.

Chart: *Prices of renewable energy technologies*

## 6. Conclusion

CO2 emission output is based on greenhouse gas but is certainly influenced by other factors. Transport, food distribution, fossil fuel consumption, population and renewable energy technologies do have an impact. With Qatar, China and USA being the top 3 of largest output. Analyses on Qatar’s CO2 output reveals increase over the past 5 years. Even though Qatar did sign the Paris Accord, no active plans for reducing are visible yet.

## 7. Appendices

Github – <https://github.com/JoKa1986/Final-Project-CO2-emission---Winc-Academy>

Analysis Report: placed inGuthub mao

Analysed datasets: placed in Github map

Colab Notebooks: placed in Github map

Colab Notebooks links:

* *CO2 output*:

<https://colab.research.google.com/drive/1Nf8757FkmgZlecEHYOoq-LbpqDr4zOD0>

* *Active Decrease CO2:*

<https://colab.research.google.com/drive/1nSxee6nOVlpVbXISOTUvPg_FnTO9oe_h>

* *Renew. Tech:*

<https://colab.research.google.com/drive/1kItj19nOdvd9AX48D2ZxFWG4a7R99ncx#scrollTo=tgKlQuEcLR3Q>

## 8. References

United Nations (UN) - Paris Accord (2015) – list of countries

Our World in Data – for datasets and reference map of countries with active plan

Matplotlib – for charts

Pandas – for analysis tools

Google Search – example reports, help with charts