Data Analysis on CO2 Emissions using Python

Introduction/Overview:

The subject of this report is to analyze the data on CO2 emissions and answer the below three questions.

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?

Data & Model:

In order to answer the questions as part of this analysis, I have performed the below steps.

- Data extraction
- Data preparation
- Visualization

Data Extraction:

I have gone through the website: https://ourworldindata.org/ to collect the relevant datasets for the analysis and identified the below datasets.

For question 1 & 2,

https://raw.githubusercontent.com/owid/co2-data/refs/heads/master/owid-co2-data.csv

For question 3,

https://raw.githubusercontent.com/owid/owid-datasets/refs/heads/master/datasets/Renewable%20energy%20costs%20(IRENA%2C%202020)/Renewable%20energy%20costs%20(IRENA%2C%202020).csv

Preparation:

To read the raw data from the csv files, I have used "pandas" library and its data frames to store the data. Performed an overall analysis of the datapoints/fields we are dealing with. Then, to prepare the data for answering the questions, I have performed the below steps,

- Renaming of columns to have shorter and appropriate column names
- Selecting only relevant data fields for the analysis
- Removed/cleaned the records which does not values for my selected fields
- Filtered the data based on the requirement to get more appropriate figures
- Calculated the measures and correlations to find derive the results

<u>Visualization:</u>

To visualize the analyzed data points, I have used "Matplotlib" library and created a graphical representation for each of the questions.

Results:

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

To provide data insights for this question, I have considered the 5 factors from the dataset to perform the correlations to the CO2 emissions per capita of a country. CO2 emissions from "Coal", "Oil", "Gas", "Cement" and "Landusechange". Filtered out the data only from 2000 to get more appropriate correlations.

1.a) In below table, it shows the top predictor per capita of a country. Every country has its own large factor for the CO2 emissions.

Country	Top Predictor
Afghanistan	Oil
Albania	Cement
Algeria	Gas
Andorra	Oil
Argentina	Oil
Armenia	Gas
Australia	Coal
Austria	Oil
Azerbaijan	Gas
Bangladesh	Oil
Belarus	Gas
Belgium	Oil
Benin	Oil
Bolivia	Oil
Bosnia and Herzegovina	Coal
Botswana	Coal
Brazil	Oil
Brunei	Coal
Bulgaria	Coal
Canada	Coal
Chile	Coal
China	Coal
Colombia	Gas
Croatia	Oil
Cuba	Oil
Cyprus	Oil
Czechia	Coal
Democratic Republic of Congo	Oil
Denmark	Coal
Dominican Republic	Coal

Egypt	Gas
El Salvador	Oil
Estonia	Coal
Fiji	Oil
Finland	Coal
France	Oil
Georgia	Gas
Germany	Coal
Greece	Coal
Guyana	Gas
Hungary	Gas
Iceland	Oil
India	Coal
Indonesia	Coal
Iran	Gas
Iraq	Oil
Ireland	Oil
Israel	Oil
Italy	Cement
Jamaica	Oil
Japan	Oil
Jordan	Oil
Kazakhstan	Coal
Kuwait	Oil
Kyrgyzstan	Oil
Latvia	Oil
Lebanon	Oil
Libya	Oil
Liechtenstein	Oil
Lithuania	Oil
Luxembourg	Oil
Malaysia	Coal
Malta	Oil
Mexico	Oil
Moldova	Oil
Morocco	Oil
Mozambique	Oil
Myanmar	Oil
Netherlands	Cement
New Zealand	Oil
Niger	Oil
Nigeria	Landusechange
Norway	Oil
Oman	Cement
Pakistan	Coal
Panama	Oil

Peru	Gas
Philippines	Coal
Poland	Coal
Portugal	Oil
Qatar	Oil
Romania	Gas
Russia	Gas
Rwanda	Oil
Saudi Arabia	Oil
Senegal	Oil
Serbia	Coal
Singapore	Oil
Slovakia	Coal
Slovenia	Coal
South Africa	Coal
South Korea	Coal
South Sudan	Oil
Spain	Oil
Sudan	Oil
Suriname	Oil
Sweden	Oil
Switzerland	Oil
Syria	Oil
Tajikistan	Cement
Tanzania	Gas
Thailand	Gas
Tunisia	Gas
Turkey	Cement
Turkmenistan	Gas
Ukraine	Coal
United Arab Emirates	Gas
United Kingdom	Oil
United States	Coal
Uruguay	Oil
Uzbekistan	Gas
Venezuela	Oil
Vietnam	Coal
Yemen	Oil

1.b) Below figure, From the top predictor of CO2 emissions per country, I have created a pie chart with the percentages to show the overall factors of CO2 across all countries. I have depicted a "red" color for the top most factor and "orange", "yellow", "blue" and "green" for the subsequent factors of CO2 emissions to show the difference.

From this dataset(from the year 2000), It shows that CO2 emissions from the "Oil" is the large factor for the overall CO2 emissions across all countries.

Distribution of Top CO₂ Contributors per Capita across all Countries
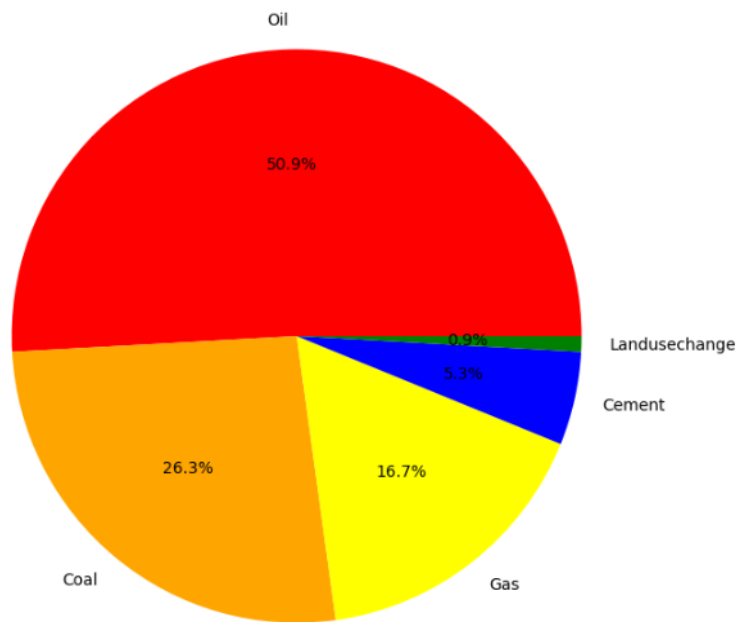
2. **Which countries are making the biggest strides in decreasing CO2 output?**

To provide data insights for this question, I have considered the data only from 2000 to get more appropriate relative change. I have calculated the latest and earliest(in this case, year 2000) CO2 emissions per capita and found the relative change.

2.a) Below figure shows the top 10 countries with the biggest strides in decreasing the overall CO2 output from the relative change. From the results, Luxembourg have taken measures to reduce CO2 emissions.

Top 10 Countries with Biggest Strides in Decreasing CO2 Output
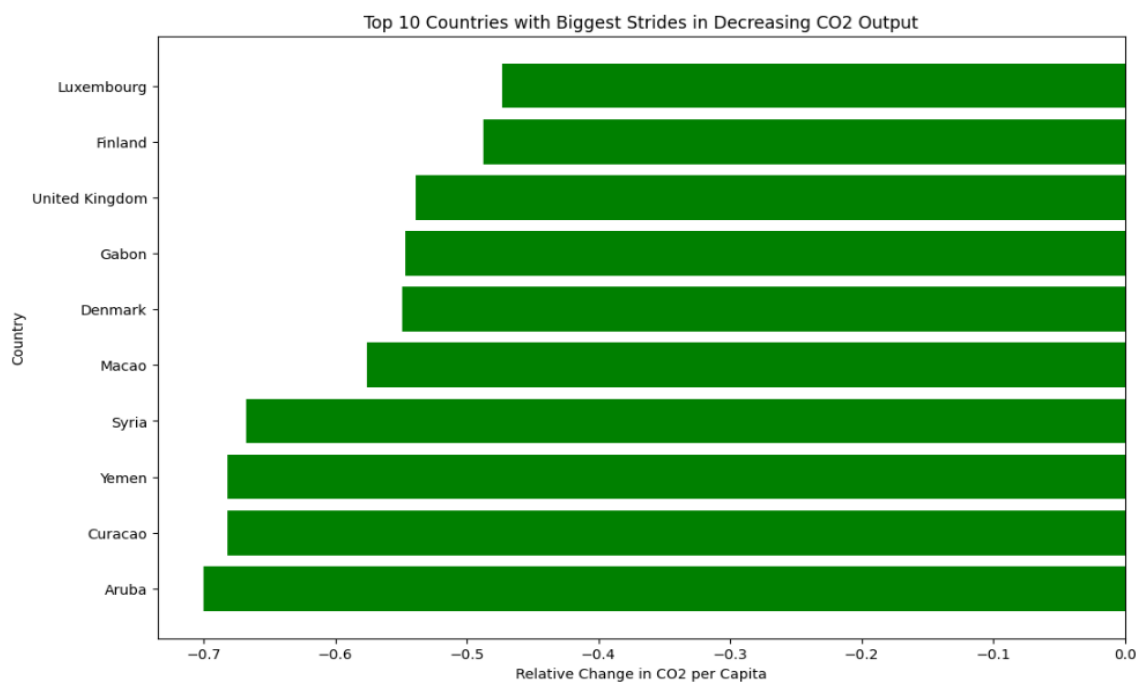
3. **Which non-fossil fuel energy technology will have the best price in the future?**

To provide data insights for this question, I have read the data from the renewable energy dataset from "ourworldinddata", unfortunately we do not have latest pricing data of various renewable energy. However, the dataset I could find was from 2019. In order to predict the future prices, I have used the "Linear regression" from the library "sklearn.linear_model". From this dataset, I have filtered "world" as an Entity because, we would like to get the best priced "non-fossil" fuel energy in the world. Since, we only have data appropriately filled in from 2010, I have used the linear regression model to predict the price until 2035 to have prediction figures more accurate(Not exactly, but may be around 70 to 80% accuracy)

3.a) Below figure is depicted to show the predicted cost until 2035 and in 2035, the best priced "Non fossil fuel energy". Result of this is, "Solar" will be the cheapest Non-fossil fuel technology in 2035 with the predicted cost of "$-0.4267" per kwh.
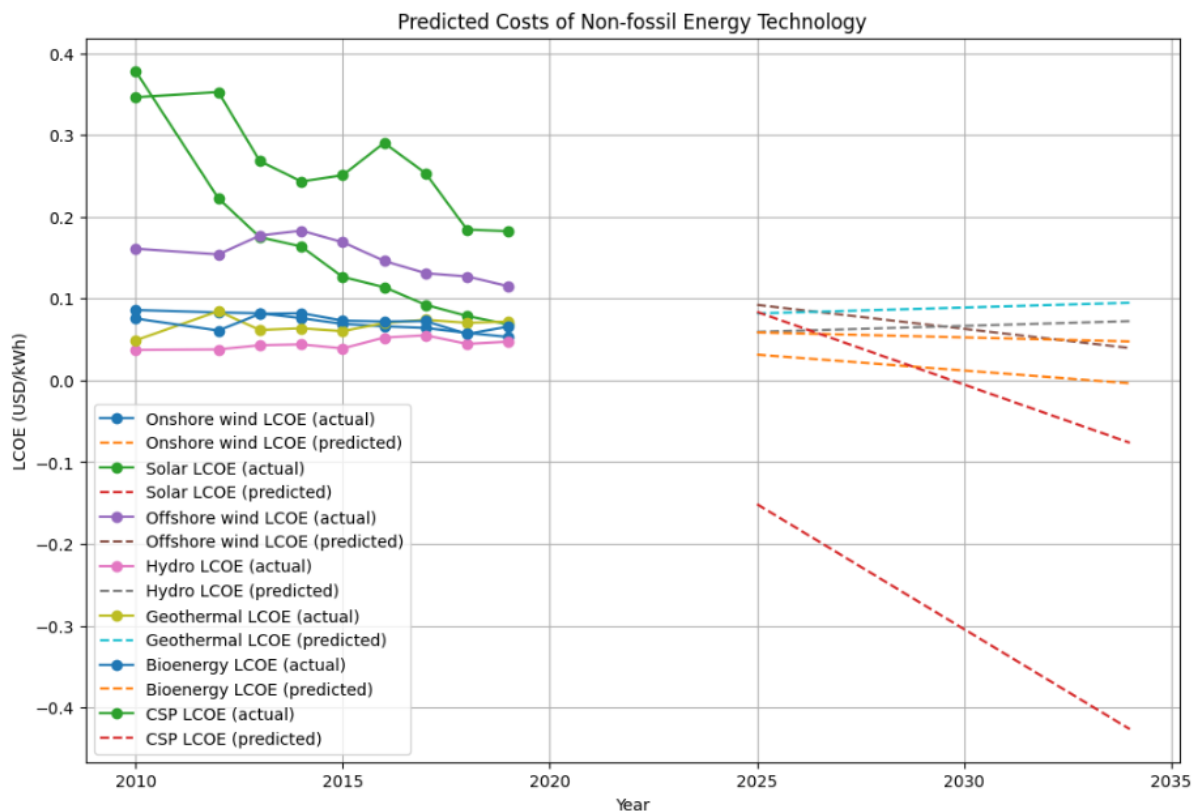
The cheapest Non-fossil energy technology in 2035 is expected to be **Solar LCOE** with a predicted cost of **$-0.4267 per kWh**.

Conclusion:

To conclude from the analysis,

1. What is the biggest predictor of a large CO2 output per capita of a country?
   Differs per country, however "Oil" is the biggest predictor considering all the countries.

2. Which countries are making the biggest strides in decreasing CO2 output?
   Considering the data from 2000, "Luxemburg" is the top most country which is making efforts to reduce the CO2 emissions.

3. Which non-fossil fuel energy technology will have the best price in the future?
   From the analysis, "Solar" will be the cheapest Non-fossil fuel technology in 2035.

Appendix:

Link to the ipython notebook file which I have created to perform this analysis.

https://github.com/priyadharshinisugumaran/Dataanalysis_P/blob/main/DA_%20CO2%20Emissions.ipynb