In this EDA I analyzed data coming from two datasets: “world_data” e “sustainable_energy”.

I realized the entire project using PgAdmin, a DBMS using SQL. Before starting, I had to put data under ‘preprocessing’, meaning transforming and re-elaborating data to make it more readable and fit to exploration. For example, I had to eliminate symbols like “%” or “$”; I had to insert all data as ‘Varchar’ and only then to modify each column.

Analysis Goals

The main purpose of this EDA is to find out which features are correlated between them and if each proves or counterproves common knowledge assumptions.

Also, I’ll delve into energy topic, consumptions and main resources in every country of the world, and the changes through the years (2000-2020).

Quick Exploration of Data and feature engineering

For a better exploration of data, I created two new features *Gdp pro capite* (“gdp_pp”) e *inquinamento pro capite* (co2_per_abitante, pollution pro capite) country-wise. Since every country have different population, I thought it was crucial to add a pro-capita metric to make more equanimous judgements.

The following are some considerable correlations found in dataset “world_data”:

- *[2.1] Physicians every 1000 inhabitants; Life Expectancy:*
  = 0,71
- *[2.2] Life Expectancy; Fertility Rate:*
  = -0.85
- *[2.3] Life Expectancy; Infant Mortality*
  = -0.92
- *[2.4] Life Expectancy; CO2 per capita:*
  = 0.45
- *[2.5] PIL pro capite; Infant Mortality*
  = -0.49
- *[2.6] Pil pro capite; Minimum Wage Amount*
  = 0.84
- *[2.7] Co2; Total Population*
  = 0.81
- *[2.8] Mean and standard deviation: CO2 per capita*
  = 0.0042 e 0.0053

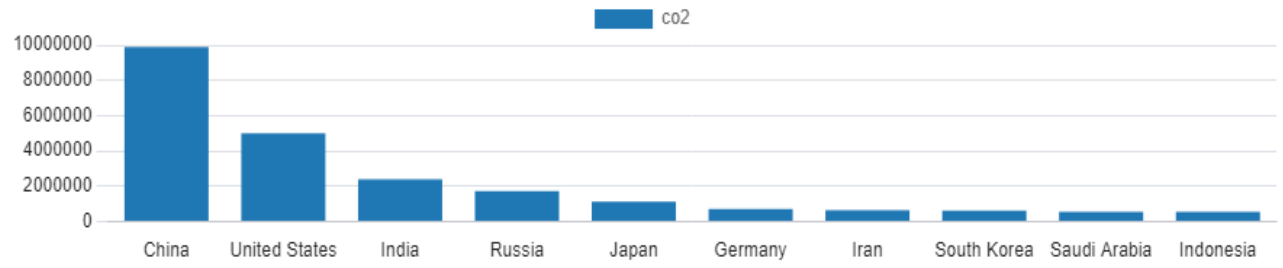
CO2 and **population** are strongly correlated, as one can easily imagine; nevertheless there are some countries with higher values of pollution, due to different industrial plans or, for example regarding China, a policy of waste disposal from other nations.

Correlation between **number of physicians** and **life expectancy [2.1]** is likely a spurious relationship, explained by a third unexplicit factor (most likely, the wealth of the country); the same goes for Life **Expectancy, Fertility Rate and Infant Mortality**: all of them are likely to depend on wealth, directly-indirectly and indirectly, respectively.

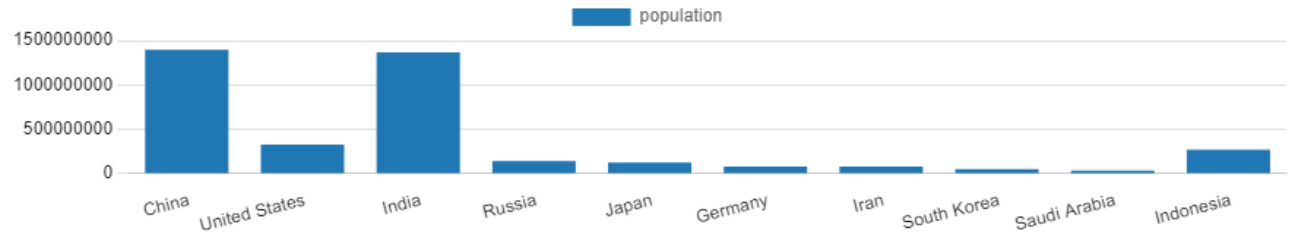
Regarding fertility rate we should although consider some cultural factors: in so called 'First World', many couples prefer to live 'childless' - called "DINKS" Double Income No Kids" – and this negatively impacts of fertility rate.

Countries' Energy

Following a graph to show first 10 countries for CO2 emissions [3]:



[CO2 in tons]

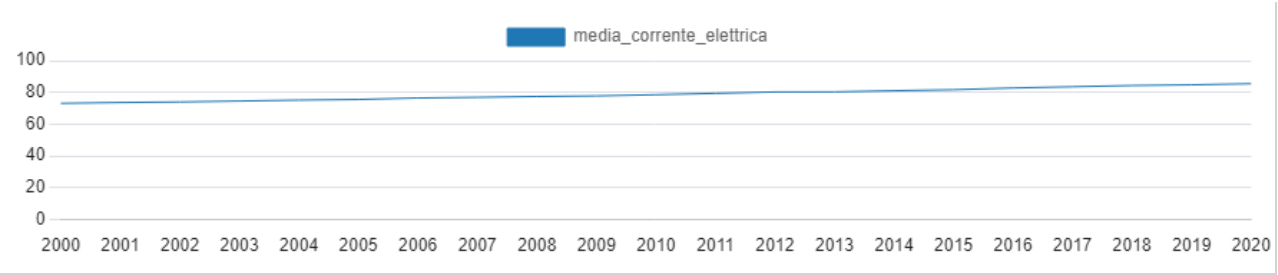


The graph above shows a discrepancy in the order of countries according to CO2 and population, specially about India and USA; these two countries allow their own population to live according to very different lifestyles and logically, gpd pro capita for USA is significantly higher. That's why GDP and CO2 are so correlated [4]: 0.92

Some data about sustainable energy

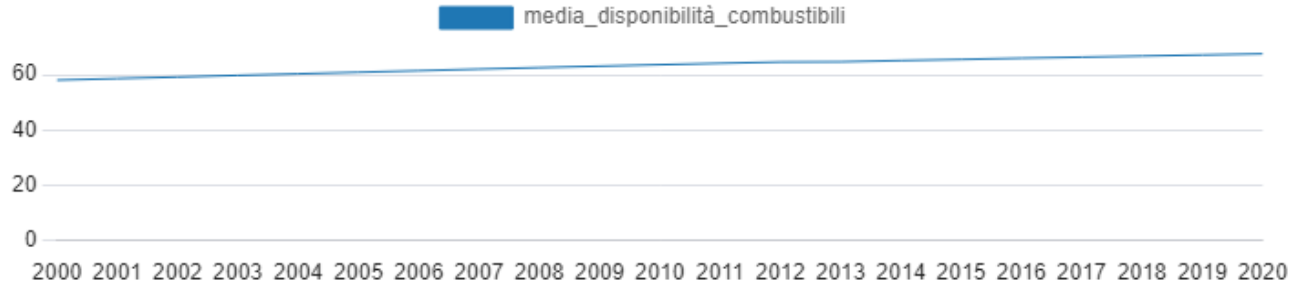
The dataset collects time-series data, which gives us a possibility to describe trends:

- *[5] Access to electricity, % of population, global mean:*



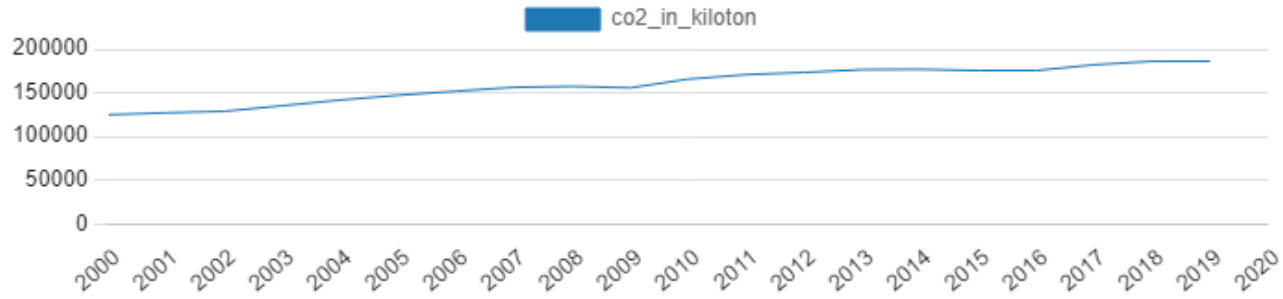
[from 73,12% to 85,46%]

- *[5.1] Access to fuels, % of population, global mean:*

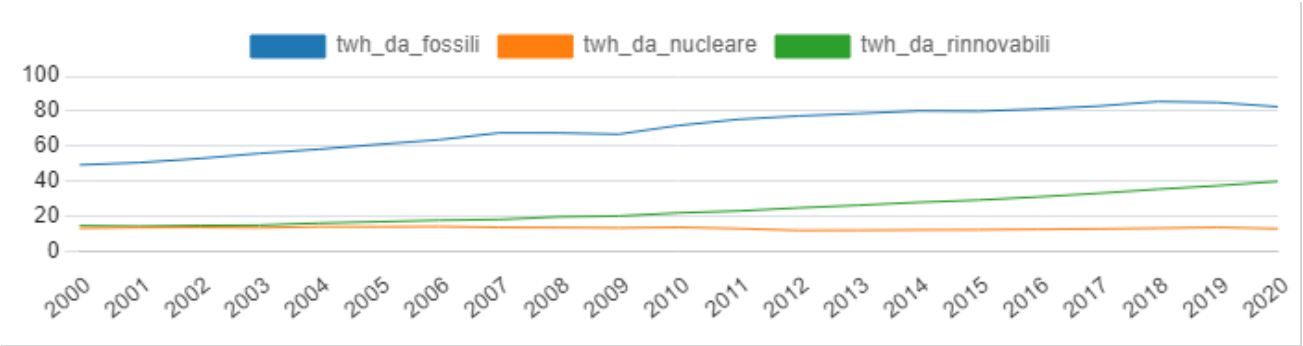


[from 58% to 67,56%]

- *[5.2] CO2 emissions in kiloton:*

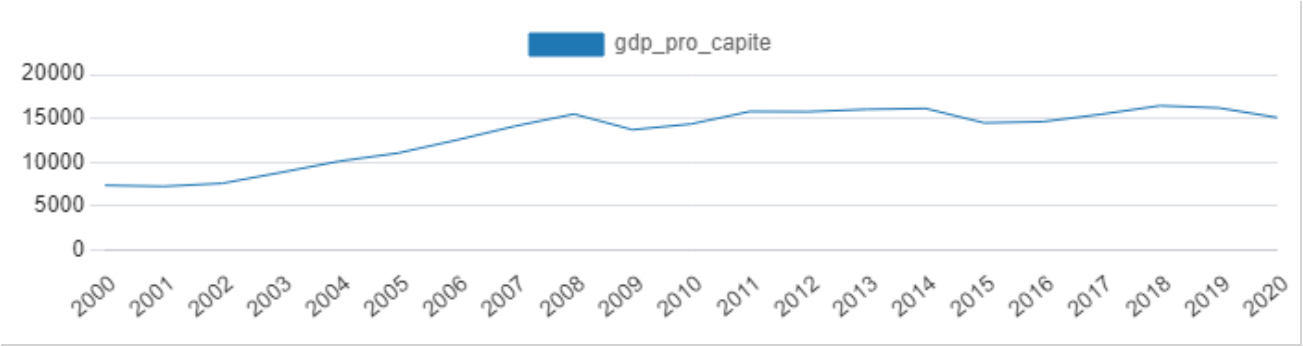


- *[5.3] Energy Sources: Renewable, Fossils, Nuclear in Twh:*



Nuclear energy is the only source that did not rise through the years, keeping itself below 20%.

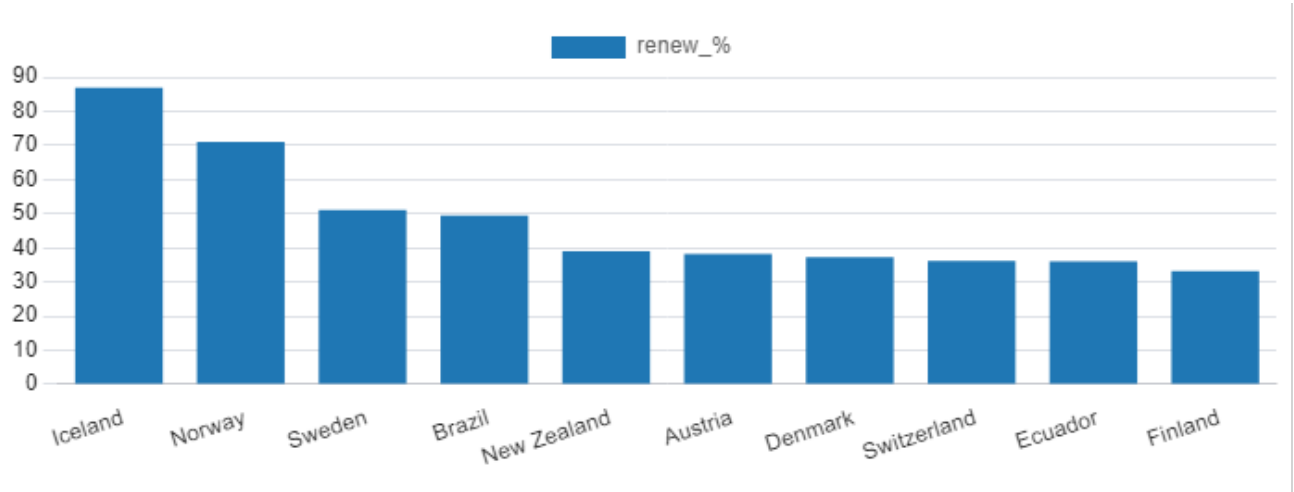
The trend for fossil energy follows closely the trend for GDP pro capita [5.4]:



[in dollars]

We can see trend declines around 2009 and 2020: they are most likely due to financial crysis of 2008 and COVID-19 in 2020.

There is of course a very high heterogeneity in renewable energy access: the following are the top 10 consumer of renewable energy as percentage of own total consumption [5.5]:



What are the countries that improved the most their access to energy sources through the last 20 years? [5.6]

	country character varying (100)	differenza_elettricità numeric
1	Afghanistan	96.08641
2	Cambodia	69.80000
3	Bhutan	68.85000
4	Solomon Islands	68.82945
5	Bangladesh	64.20000
6	Nepal	60.85373
7	Eswatini	59.53474
8	Kenya	56.30677
9	Papua New Guinea	52.82299
10	Somalia	47.83837

Global Average: **13.27%** [5.7]

- Top 10 Countries for improving in fuel access, % of population: [5.8]

	country character varying (100)	differenza_combustibili numeric
1	Indonesia	78.40000
2	Tuvalu	54.50000
3	Bhutan	52.40000
4	Tonga	48.20000
5	Sudan	47.10000
6	Maldives	46.10000
7	Tajikistan	45.90000
8	India	45.60000
9	Guyana	45.00000
10	Albania	43.10000

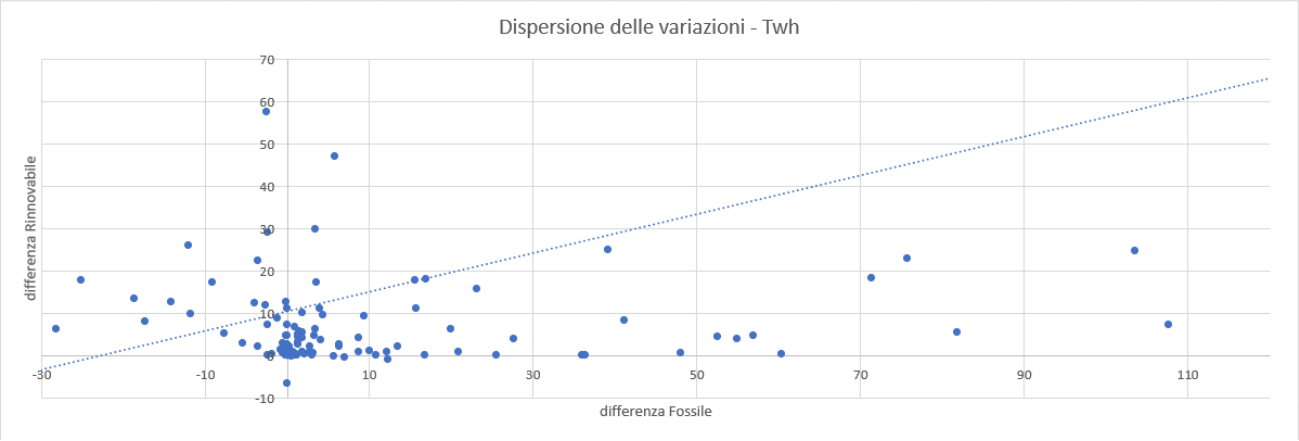
When calculating the global average growth rate and its standard deviation [5.9], a result of **9.87%** was obtained for the first and **14.89** for the second. The variance is very high, given the already advanced state of many world economies at the beginning of the 2000s or even the worsening of this condition in some nations.

- Least 10 countries for improvement in fuel access:

	country character varying (100)	differenza_combustibili numeric
1	Senegal	-10.60000
2	Vanuatu	-8.30000
3	Bosnia and Herzegovina	-6.90000
4	Zambia	-3.40000
5	Grenada	-3.30000
6	Zimbabwe	-3.10000
7	Malaysia	-2.50000
8	Gambia	-2.00000
9	Saint Vincent and the Grenadines	-1.40000
10	Malawi	-1.10000
11	Uganda	-0.40000

- ***Analysis of changes in energy inputs: comparison between renewable and fossil:***

First, similar to what was described earlier, a query was executed to obtain a summary table of the differences between the year 2020 and the year 2000 in the use of renewable and fossil energy, using as reference the columns containing data in TWh (terawatt-hours) and grouping by nations. Using the previous one as a CTE, I calculated the correlation between the two variables, which turned out to be 0.93. This means that consumption has globally increased for both energy sources (as also seen on page 5) and that renewable sources have not replaced fossil ones in the energy supply mix.



[made with Excel; excluded some outliers]

- *"Energetic Virtuosity" calculated as the increase in consumption derived from fossil fuels but offset by consumption from renewables, values in TWh [6]:*

	country character varying (100)	differenza_fossile numeric	differenza_rinnovabile numeric	virtuosismo_energetico numeric
1	China	4070.83	1959.38	-2111.45
2	India	726.99	235.49	-491.50
3	Saudi Arabia	199.14	0.21	-198.93
4	Indonesia	160.48	33.31	-127.17
5	Egypt	107.69	7.13	-100.56
6	Mexico	103.66	24.68	-78.98
7	United Arab Emirates	81.95	5.50	-76.45
8	Bangladesh	60.48	0.37	-60.11
9	Thailand	71.37	18.35	-53.02
10	Malaysia	75.86	22.98	-52.88

[10 most impactful]

	country character varying (100)	differenza_fossile numeric	differenza_rinnovabile numeric	virtuosismo_energetico numeric
1	United States	-265.38	470.47	735.85
2	Germany	-115.82	216.01	331.83
3	United Kingdom	-154.56	121.76	276.32
4	Brazil	52.28	211.24	158.96
5	Italy	-57.11	66.03	123.14
6	Canada	-53.37	65.54	118.91
7	Spain	-36.58	79.30	115.88
8	France	-2.47	57.45	59.92
9	Denmark	-25.15	17.88	43.03
10	Australia	5.87	46.88	41.01

[10 least impactful]

These differences are clearly also due to the starting point of the states mentioned above: an economy that was already advanced in 2000 had less need to increase global consumption and could experiment more in the field of innovative energies. The average growth of GDP per capita is added to the same previous query by performing an inner join) [6.1]:

	country character varying (100)	differenza_fossile numeric	differenza_rinnovabile numeric	virtuosismo_energetico numeric	crescita_media_gdp numeric
1	Barbados	0.17	0.07	-0.10	-0.24
2	Greece	-18.67	13.41	32.08	-0.11
3	Italy	-57.11	66.03	123.14	-0.04
4	Zimbabwe	0.10	0.75	0.65	0.13
5	Jamaica	-2.33	0.18	2.51	0.30
6	Portugal	-9.08	17.24	26.32	0.47
7	Japan	138.38	101.44	-36.94	0.60
8	Dominica	0.08	0.01	-0.07	0.74
9	France	-2.47	57.45	59.92	1.00
10	Germany	-115.82	216.01	331.83	1.09

[Lowest GDP Growth]

Cambodia	3.41	4.57	1.16	7.22
Somalia	0.12	0.03	-0.09	7.37
Tajikistan	1.41	4.34	2.93	7.54
Turkmenistan	10.86	0.00	-10.86	8.07
Azerbaijan	7.03	-0.30	-7.33	8.15
Equatorial Guinea	0.84	0.48	-0.36	8.26
Qatar	36.04	0.02	-36.02	8.43
China	4070.83	1959.38	-2111.45	8.70
Ethiopia	-0.05	12.58	12.63	8.80
Myanmar	9.45	9.20	-0.25	9.53

[Highest GDP Growth]

Some of the countries with the lowest GDP growth are also among the countries that have most redirected their supplies toward "green" energy (Italy, Germany, France) and vice versa, but from the data, it cannot be stated that renewable energies are replacing fossil fuels globally.

- *[6.2] Decarbonization as a percentage of the starting point, meaning relative increases and decreases in the use of coal for electricity production:*

	country character varying (100)	decarbonizzazione numeric
1	Congo	-74.93
2	Ghana	-55.34
3	Haiti	-40.87
4	Rwanda	-35.16
5	Madagascar	-34.71
6	Cameroon	-33.10
7	Togo	-32.96
8	Burundi	-31.43
9	Mali	-27.85
10	Sao Tome and Principe	-23.33

[Top 10 for decreasing]

	country character varying (100)	decarbonizzazione numeric
1	Sierra Leone	77.98
2	Denmark	66.15
3	Liberia	59.55
4	Estonia	48.73
5	Cambodia	44.03
6	Kenya	42.05
7	Luxembourg	39.51
8	Ireland	37.11
9	Nicaragua	34.09
10	United Kingdom	34.01

[Top 10 for increasing]

- *[6.3] Which nations have most significantly changed their use of renewable energy as a percentage of the total?*

	country character varying (100)	ricorso_rinnovabili numeric
1	Philippines	-7.56
2	Kazakhstan	-2.12
3	Peru	-1.76
4	Sri Lanka	-1.75
5	Azerbaijan	-1.47
6	Norway	-1.44
7	Iraq	-1.28
8	Bangladesh	-0.97
9	Egypt	-0.86
10	Uzbekistan	-0.38

	country character varying (100)	ricorso_rinnovabili numeric
1	Denmark	30.17
2	Iceland	19.03
3	United Kingdom	18.95
4	Portugal	18.65
5	Ireland	18.51
6	Germany	18.23
7	Spain	15.09
8	Ecuador	14.43
9	Estonia	14.04
10	Finland	13.47

- *[6.4Percentage and TWh variation in renewable energy production (2000-2020) for each analyzed country:*

	country character varying (100)	var_ricorso_rinn_% text	twh_2020 numeric (10,2)	var_rinn_twh numeric
1	Philippines	-7.56%	23.21	3.84
2	Kazakhstan	-2.12%	11.94	4.41
3	Peru	-1.76%	33.72	17.65
4	Sri Lanka	-1.75%	5.94	2.77
5	Azerbaijan	-1.47%	1.22	-0.30
6	Norway	-1.44%	152.09	9.98
7	Iraq	-1.28%	5.30	4.69
8	Bangladesh	-0.97%	1.31	0.37
9	Egypt	-0.86%	20.83	7.13
10	Uzbekistan	-0.38%	5.01	-0.81

[Worst 10]

	country character varying (100)	var_ricorso_rinn_% text	twh_2020 numeric (10,2)	var_rinn_twh numeric
1	Denmark	30.17%	23.45	17.88
2	Iceland	19.03%	18.78	11.17
3	United Kingdom	18.95%	131.74	121.76
4	Portugal	18.65%	30.11	17.24
5	Ireland	18.51%	13.46	12.28
6	Germany	18.23%	251.48	216.01
7	Spain	15.09%	113.79	79.30
8	Ecuador	14.43%	24.83	17.30
9	Estonia	14.04%	2.97	2.96
10	Finland	13.47%	35.93	12.55

[Top 10 – mostly N.Europe]

- *[6.5] Correlation between changes in the use of renewable energy and decarbonization:* **= 0.70**

- *[6.6] European countries TWh production from renewable energy:*

	country character varying (100)	twh_renew numeric (10,2)
1	Germany	251.48
2	France	125.28
3	Italy	116.90
4	Spain	113.79
5	Austria	55.42
6	Finland	35.93
7	Portugal	30.11
8	Belgium	23.46
9	Greece	17.55
10	Slovakia	6.85

- *[6.7] Changes in energy consumption per capita, 2000-2020*

	country character varying (100)	variazione_kwh_pp numeric
1	Malta	25687.41
2	Latvia	5755.75
3	Estonia	4496.24
4	Lithuania	3246.35
5	Portugal	-2694.05
6	Slovenia	-3080.54
7	Austria	-3835.01
8	Spain	-7104.41
9	Slovakia	-7397.96
10	Greece	-7453.21
11	Germany	-7526.02
12	Cyprus	-9060.13
13	Italy	-9087.49
14	Finland	-12740.22
15	France	-14903.18
16	Belgium	-14986.68
17	Luxembourg	-21160.64

[2020-2000]

	country character varying (100)	variazione_kwh_pp numeric
1	Malta	33703.46
2	Latvia	7614.72
3	Estonia	7573.30
4	Lithuania	3194.73
5	Portugal	-7.57
6	Austria	-549.02
7	Slovenia	-1717.49
8	Spain	-3730.84
9	Greece	-3830.42
10	Germany	-4435.95
11	Cyprus	-5470.56
12	Italy	-6387.23
13	Slovakia	-6698.51
14	Belgium	-8268.15
15	Luxembourg	-9128.73
16	Finland	-10157.58
17	France	-10639.80

[2019-2000, considering changes due to COVID]

Final Considerations

With this analysis, the global scenario regarding energy consumption and changes in its composition across different nations of the world has been explored.

Regarding the variations observed over these 20 years, it's important to highlight two aspects:

- At the beginning of the reference period (2000), the international balance was different from how it is today, and many of the observed nations were in a significantly 'worse' status compared to their current one. For these countries, therefore, it is natural to observe growth: indeed, for a country already established in the international context, it is difficult to grow further, and it is more likely that it will try to maintain its position, or increase it with contained variations.
- Due to this heterogeneity in the starting points of different countries, it will be interesting to propose a similar analysis in the not necessarily immediate future (therefore, at least 2030): in this way, it will be possible to observe how today's leading countries have behaved in relation to the environmental challenges that the future holds for humanity.

The constant increase in world population and, consequently, the increase in consumption, will force all governments to face choices of an energy nature with consequences both on an economic and ethical level.

To date, international cooperation regarding environmental and ecological sustainability has not yielded the fruits that many hope for: there are great historical differences between the various countries of the world, and today's emerging countries understandably have desires for expansion and redemption from their past.

Humanity's greatest challenge in the future may not necessarily be technological in nature, but rather diplomatic.