CareerFoundry – Data Analytics

Achievement 6 – Is Brazil really a climate killer?

Mariana Oliveira, 28.07.2024

Objective

Analyze key metrics of emission, energy and economy, showcasing the numbers and facts to provide a more complete picture of Brazil in the climate scenario.

Context

More and more are discussions about climate change not only a topic for politics, but also between friends and family. As a Brazilian living in Germany for some years, I have the impression that there are some misconceptions about the role of Brazil in the climate change. I would like to present the data about it to show that Brazil is not only “burning the Amazonian forest”, but actually performs quite well in terms of renewable energy and forest coverage.

It is also important to point the climate performance and economic development are tied together. For example, countries where the population do not have full electricity coverage will also likely to have lower greenhouse emissions. But it is an option to leave them out of technological advancements because they are “green enough”?

The idea is not to provide a black and white picture of the countries, but to show that there are different “shades of green” on the topic.

Figure 1: Forest in Fire or Renewable Energy - Designed by Freepik

When you think of news about Brazil, what image do you associate with it the most?

Key Questions

For this analysis I will break down the key indices of Brazil comparing it worldwide and also with only a selection of countries:

- How are the CO_2 emissions per capita in Brazil compared to the rest of the world?
- How is the forest coverage in Brazil and other countries?
- How is the deforestation rate in the last years?
- What is the share of renewable energy worldwide?
- What is the specific share on renewable sources on electricity generation?

How does the development of a land relate to the emissions?

Data Choice

In order to answer the questions proposed for this project, I looked for historical and global data covering following topics:

- CO_2 Emission – being one driving factor to the global warming and climate change
- Renewable Energy (and share in electricity) – energy production is one of the
- Forest Area (and deforestation) – as measure of the forest protection: how much still exists and
- (Inequality-adjusted) Human Development Index – I find this index more appropriate than the "usual" Human Development Index since it accounts for economical inequalities within a country.

Data Source

All data showcased in this project was extracted from, *Our World in Data*, which is a project of *Global Change Data Lab*, a non-profit organization in the United Kingdom.

- https://ourworldindata.org/about
- https://global-change-data-lab.org/

This is a very interesting data source, which provides not only reliable data for free use, but as well reports and analysis on different topics.

To provide a broader picture for my analysis, I used different datasets which were complimentary. Each dataset was extracted from a different topic and is acknowledged individually:

CO_2 Emissions (MAIN):

- Hannah Ritchie, Pablo Rosado and Max Roser (2023) - "CO_2 and Greenhouse Gas Emissions" Published online at OurWorldInData.org. Retrieved from: 'https://ourworldindata.org/co2-and-greenhouse-gas-emissions' [Online Resource]

Renewable Energy and Renewable Electricity

- Hannah Ritchie, Max Roser and Pablo Rosado (2020) - "Renewable Energy" Published online at OurWorldInData.org. Retrieved from: 'https://ourworldindata.org/renewable-energy' [Online Resource]

Forest Area and Deforestation

- Hannah Ritchie, Fiona Spooner and Max Roser (2021) - "Forests and Deforestation" Published online at OurWorldInData.org. Retrieved from: 'https://ourworldindata.org/forests-and-deforestation' [Online Resource]

(Inequality-adjusted) Human Development Index

- Bastian Herre and Pablo Arriagada (2023) - "The Human Development Index and related indices: what they are and what we can learn from them" Published online at OurWorldInData.org. Retrieved from: 'https://ourworldindata.org/human-develoepment-index' [Online Resource]

If I find necessary to include more datasets later during the analysis, I will also source then on *Our World in Data*, an acknowledge it properly on the final report.

Project Deliverables

- GitHub Repository with script and data
- Storyboard in Tableau Public
- CSV files of the summarized data (available in GitHub)
- Short overview of results (available in GitHub)

Data Profile

Limitations and Ethics

The data is open to use, according to the source Our World in Data:

https://ourworldindata.org/faqs#can-i-use-or-reproduce-your-data

> *" All data produced by Our World in Data is completely open access under the Creative Commons BY license. You have the permission to use, reproduce, and distribute it in any medium, provided the source and authors are credited.*
>
> *The data produced by third parties and made available by Our World in Data is subject to the license terms from the original third-party authors. We will always indicate the original source of the data in our documentation, so you should always check the license of any such third-party data before use and redistribution."*
>
> *Our World in Data*

Since the data does not contain personal data, I can rest sure that there will not be privacy concerns.

However, there might be cultural bias in the analysis. Since I also analyze the development index, I will pay attention to the choice of wording as not to refer to a country as "poor", "underdeveloped" and other adjectives that could be interpreted as depreciating.

Data Wrangling

I believe it is interesting to point out that the original dataset were really heterogenous as to how many countries (or entities) were considered and also the time span each one covered.

Dataset name	*Number of countries*	*Begins in*	*Ends in*
CO2 Emissions	261	1750	2022
Human Development Index	178	2010	2022
Forest area	263	1000	2020
Deforestation	263	1000	2020
Renewable energy	100	1965	2023
Renewable electricity	250	1985	2023

The countries of interest were chosen trying to be spread worldwide, but in a continent being geographically close but having different inequality adjusted Human Development Index (here abbreviated as iaHDI, data from 2022):

Continent	*Country 1*	*iaHDI*	*Country 2*	*iaHDI*
South America	Brazil	0,577	Argentina	0,747
North America	United States of America	0,823	Mexico	0,641
Europe	Germany	0,881	Portugal	0,774
Africa	Egypt	0,561	Sudan	0,331
Asia – east	China	0,662	Japan	0,844
Asia – west	Oman	0,721	Iran	0,584
Oceania	Australia	0,860	Papua New Guinea	0,407

The original datasets contained much more data than necessary. In order to spare work and computing time. I have selected only the columns, countries and time range of interest.

Dataset name	*Data frame abbreviation*	*Original size*		*After wrangling*	
		Rows	Columns	Rows	Columns
CO2 Emissions	co2	47415	79	2453	79
Human Development Index	hdi	2106	4	144	4
Forest area	forest	7974	4	152	4
Deforestation	deforest	495	4	16	4
Renewable energy	energy	4879	4	121	4
Renewable electricity	elect	7152	4	154	4

Data Cleaning

The whole dataset consists of big data, regarding CO2 (and other) emissions.

I have already taken the countries and time of interest during the data wrangling.

The only two issues were

- Many values missing in the CO2 dataset. I kept them at first since they do not appear to interfere with the most interesting values
- One column in CO2 dataset, 'iso code', had mixed type values and I converted them all to strings

Data Understanding

To have a second impression, I took out the main information and basic statistics of the cleaned data, which is the one I will be working with from now on.

Dataset CO2 Emissions

```
<class 'pandas.core.frame.DataFrame'>
Index: 2453 entries, 160 to 47412
Data columns (total 79 columns):
 #   Column                                Non-Null Count  Dtype
---  ------                                --------------  -----
 0   Country                               2453 non-null   object
 1   Year                                  2453 non-null   int64
 2   iso_code                              2453 non-null   object
 3   population                            2376 non-null   float64
 4   gdp                                   1804 non-null   float64
 5   cement_co2                            2298 non-null   float64
 6   cement_co2_per_capita                 2298 non-null   float64
 7   co2                                   2354 non-null   float64
 8   co2_growth_abs                        2354 non-null   float64
 9   co2_growth_prct                       2354 non-null   float64
 10  co2_including_luc                     2134 non-null   float64
 11  co2_including_luc_growth_abs          2134 non-null   float64
 12  co2_including_luc_growth_prct         2134 non-null   float64
 13  co2_including_luc_per_capita          2134 non-null   float64
 14  co2_including_luc_per_gdp             1782 non-null   float64
 15  co2_including_luc_per_unit_energy     2077 non-null   float64
 16  co2_per_capita                        2354 non-null   float64
 17  co2_per_gdp                           1804 non-null   float64
 18  co2_per_unit_energy                   2242 non-null   float64
 19  coal_co2                              2354 non-null   float64
 20  coal_co2_per_capita                   2354 non-null   float64
 21  consumption_co2                       1317 non-null   float64
 22  consumption_co2_per_capita            1317 non-null   float64
 23  consumption_co2_per_gdp               1306 non-null   float64
 24  cumulative_cement_co2                 2298 non-null   float64
 25  cumulative_co2                        2354 non-null   float64
 26  cumulative_co2_including_luc          2134 non-null   float64
 27  cumulative_coal_co2                   2354 non-null   float64
 28  cumulative_flaring_co2                2354 non-null   float64
 29  cumulative_gas_co2                    2354 non-null   float64
 30  cumulative_luc_co2                    2167 non-null   float64
 31  cumulative_oil_co2                    2354 non-null   float64
```

```
 32   cumulative_other_co2                      506 non-null    float64
 33   energy_per_capita                         2253 non-null   float64
 34   energy_per_gdp                            1804 non-null   float64
 35   flaring_co2                               2354 non-null   float64
 36   flaring_co2_per_capita                    2354 non-null   float64
 37   gas_co2                                   2354 non-null   float64
 38   gas_co2_per_capita                        2354 non-null   float64
 39   ghg_excluding_lucf_per_capita             2123 non-null   float64
 40   ghg_per_capita                            2123 non-null   float64
 41   land_use_change_co2                       2167 non-null   float64
 42   land_use_change_co2_per_capita            2134 non-null   float64
 43   methane                                   2123 non-null   float64
 44   methane_per_capita                        2123 non-null   float64
 45   nitrous_oxide                             2123 non-null   float64
 46   nitrous_oxide_per_capita                  2123 non-null   float64
 47   oil_co2                                   2354 non-null   float64
 48   oil_co2_per_capita                        2354 non-null   float64
 49   other_co2_per_capita                      506 non-null    float64
 50   other_industry_co2                        506 non-null    float64
 51   primary_energy_consumption                2253 non-null   float64
 52   share_global_cement_co2                   2298 non-null   float64
 53   share_global_co2                          2354 non-null   float64
 54   share_global_co2_including_luc            2134 non-null   float64
 55   share_global_coal_co2                     2354 non-null   float64
 56   share_global_cumulative_cement_co2        2298 non-null   float64
 57   share_global_cumulative_co2               2354 non-null   float64
 58   share_global_cumulative_co2_including_luc 2134 non-null   float64
 59   share_global_cumulative_coal_co2          2354 non-null   float64
 60   share_global_cumulative_flaring_co2       2354 non-null   float64
 61   share_global_cumulative_gas_co2           2354 non-null   float64
 62   share_global_cumulative_luc_co2           2167 non-null   float64
 63   share_global_cumulative_oil_co2           2354 non-null   float64
 64   share_global_cumulative_other_co2         506 non-null    float64
 65   share_global_flaring_co2                  2354 non-null   float64
 66   share_global_gas_co2                      2354 non-null   float64
 67   share_global_luc_co2                      2167 non-null   float64
 68   share_global_oil_co2                      2354 non-null   float64
 69   share_global_other_co2                    506 non-null    float64
 70   share_of_temperature_change_from_ghg      2420 non-null   float64
 71   temperature_change_from_ch4               2211 non-null   float64
 72   temperature_change_from_co2               2420 non-null   float64
 73   temperature_change_from_ghg               2420 non-null   float64
 74   temperature_change_from_n2o               2211 non-null   float64
 75   total_ghg                                 2123 non-null   float64
 76   total_ghg_excluding_lucf                  2123 non-null   float64
 77   trade_co2                                 1317 non-null   float64
 78   trade_co2_share                           1317 non-null   float64
dtypes: float64(76), int64(1), object(2)
memory usage: 1.5+ MB
```

	Year	population	gdp	cement_co2	cement_co2_per_capita	co2	co2_growth_abs	co2_growth_prct	co2_including_luc
count	2453.000000	2.376000e+03	1.804000e+03	2298.000000	2298.000000	2354.000000	2354.000000	2354.000000	2134.000000
mean	2015.000000	3.432790e+07	6.515238e+11	7.039644	0.111822	160.016285	1.559339	2.367602	205.383900
std	3.162922	1.350162e+08	2.129986e+12	52.925443	0.151380	799.088814	33.816524	12.588857	877.114014
min	2010.000000	1.833000e+03	5.444287e+08	0.000000	0.000000	0.004000	-547.517000	-52.657000	-7.961000
25%	2012.000000	7.610282e+05	2.756526e+10	0.000000	0.000000	1.260250	-0.290250	-3.126500	5.372250
50%	2015.000000	6.275342e+06	9.052598e+10	0.469000	0.064000	8.658000	0.011000	1.089500	26.757500
75%	2018.000000	2.337877e+07	4.184048e+11	2.096750	0.160750	55.678500	0.692750	6.814000	88.073500
max	2020.000000	1.424930e+09	2.415184e+13	858.233000	1.004000	10914.012000	911.782000	141.744000	11743.429000

8 rows × 77 columns

Dataset HDI

```
<class 'pandas.core.frame.DataFrame'>
Index: 144 entries, 65 to 1995
Data columns (total 4 columns):
 #   Column                                    Non-Null Count  Dtype  
---  ------                                    --------------  -----  
 0   Country                                   144 non-null    object 
 1   Code                                      144 non-null    object 
 2   Year                                      144 non-null    int64  
 3   Inequality-adjusted Human Development Index  144 non-null    float64
dtypes: float64(1), int64(1), object(2)
memory usage: 5.6+ KB
```

	YEAR	INEQUALITY-ADJUSTED HUMAN DEVELOPMENT INDEX
COUNT	144.000000	144.000000
MEAN	2015.222222	0.662500
STD	3.132158	0.159494
MIN	2010.000000	0.298000
25%	2013.000000	0.574500
50%	2015.000000	0.662500
75%	2018.000000	0.812500
MAX	2020.000000	0.879000

Dataset Forest

```
<class 'pandas.core.frame.DataFrame'>
Index: 152 entries, 330 to 7475
Data columns (total 4 columns):
 #   Column         Non-Null Count  Dtype  
---  ------         --------------  -----  
 0   Country        152 non-null    object 
 1   Code           152 non-null    object 
 2   Year           152 non-null    int64  
 3   Forest cover   152 non-null    float64
dtypes: float64(1), int64(1), object(2)
memory usage: 5.9+ KB
```

	YEAR	FOREST COVER
COUNT	152.000000	152.000000
MEAN	2015.059211	29.690373
STD	3.150178	24.413098
MIN	2010.000000	0.008078
25%	2012.000000	10.365027
50%	2015.000000	32.658786
75%	2018.000000	36.154995
MAX	2020.000000	79.889779

Dataset Deforest

```
<class 'pandas.core.frame.DataFrame'>
Index: 16 entries, 10 to 466
Data columns (total 4 columns):
 #   Column         Non-Null Count  Dtype  
---  ------         --------------  -----  
 0   Country        16 non-null     object 
 1   Code           16 non-null     object 
 2   Year           16 non-null     int64  
 3   Deforestation  16 non-null     float64
dtypes: float64(1), int64(1), object(2)
memory usage: 640.0+ bytes
```

	YEAR	DEFORESTATION
COUNT	16.000000	1.600000e+01
MEAN	2012.187500	3.487606e+05
STD	2.561738	5.739312e+05
MIN	2010.000000	0.000000e+00
25%	2010.000000	3.287250e+04
50%	2010.000000	1.503100e+05
75%	2015.000000	2.804925e+05
MAX	2015.000000	1.867800e+06

Dataset Energy

```
<class 'pandas.core.frame.DataFrame'>
Index: 121 entries, 222 to 4557
Data columns (total 4 columns):
 #   Column                                Non-Null Count  Dtype  
---  ------                                --------------  -----  
 0   Country                               121 non-null    object 
 1   Code                                  121 non-null    object 
 2   Year                                  121 non-null    int64  
 3   Renewables (% equivalent primary energy)  121 non-null    float64
dtypes: float64(1), int64(1), object(2)
memory usage: 4.7+ KB
```

	YEAR	RENEWABLES (% EQUIVALENT PRIMARY ENERGY)
COUNT	121.000000	121.000000
MEAN	2015.000000	12.826171
STD	3.175426	11.715880
MIN	2010.000000	0.926311
25%	2012.000000	5.840176
50%	2015.000000	8.875714
75%	2018.000000	13.335284
MAX	2020.000000	48.693333

Dataset Electricity

```
<class 'pandas.core.frame.DataFrame'>
Index: 154 entries, 294 to 6771
Data columns (total 4 columns):
 #   Column                       Non-Null Count  Dtype  
---  ------                       --------------  -----  
 0   Country                      154 non-null    object 
 1   Code                         154 non-null    object 
 2   Year                         154 non-null    int64  
 3   Renewables - % electricity  154 non-null    float64
dtypes: float64(1), int64(1), object(2)
memory usage: 6.0+ KB
```

	YEAR	RENEWABLES - % ELECTRICITY
COUNT	154.000000	154.000000
MEAN	2015.000000	27.497394
STD	3.172595	23.383107
MIN	2010.000000	0.000000
25%	2012.000000	11.977006
50%	2015.000000	20.085413
75%	2018.000000	35.234106
MAX	2020.000000	87.193470