**Group 17 – Project 1 Write Up**

This analysis was conducted by Lina Blanco Restrepo, Herbert Dennis, Amar Patil, and

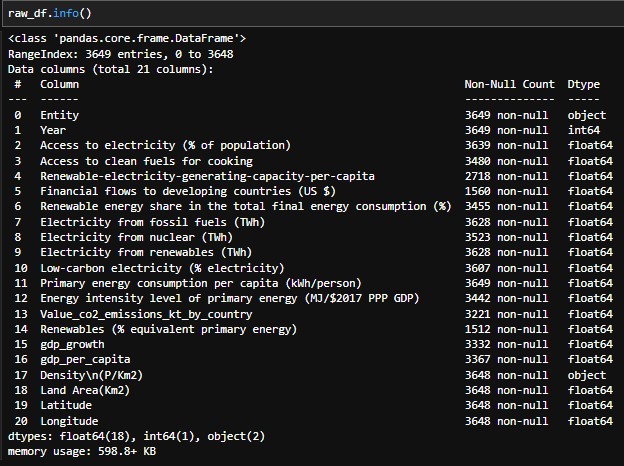
Jason Wloszek. In it we analyze sustainability data from the Kaggle Dataset, Global Data on Sustainable Energy (2000-2020)[[1]](#footnote-1). Our goal with this dataset was to analyze it and address the below research questions that we posed to determine if there was any predictability in a country’s adoption and use of sustainable energies in lieu of fossil fuel sources depending on the financial backing they receive. Highlighting the global relationship between countries in the chosen dataset, and their financial dynamics surrounding the implementation of policies for a more energy sustainable world. We will also consider extenuating factors outside our dataset, though these were less individually quantifiable due to the limitations of the dataset.

We chose this dataset because of the increased public and global focus on sustainable development. “According to the UN the rate of which the world is advancing towards sustainable energy targets – is not fast enough. At the current pace, about 660 million people will still lack access to electricity and close to 2 billion will still rely on polluting fuels and technologies for cooking by 2030.”[[2]](#footnote-2)2 This quote helps us focus the purpose behind this dataset, is the world progressing quickly enough in sustainable development? If not, what pace are we going at? What factors are going to have an impact on the progression of sustainable development? These general questions We developed higher level questions to help provide more detailed answers.

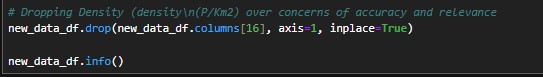
We chose 3 high level questions that we would use to try and address this issue.

1. What impact does foreign aid have in developing countries regarding renewables?
2. Are there any notable factors that promote the adoption of renewable energy year over year?
3. What conclusions can be drawn by examining a country’s renewable usage by geographical location?

To proceed with answering these questions, we first needed to look at the dataset we are working with and modify it to better fit the questions that we are posing. Below is our raw data set, it has 21 columns and provides a good working amount of data for each country within the data set over the twenty years that it covers. Reviewing the columns and their titles, we considered which columns would have the most impact in answering our questions, which columns need to be manipulated, and which can be dropped all together.

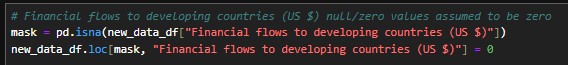


Firstly, we dropped the “Renewables (% equivalent primary energy)” and “Population – Density\n(P/Km2)” columns. The reasoning for this was to simplify the data set we were working with and to remove columns that we would not be using or presented to much clutter in the dataset. The Renewables column was something that we did not find value in, being given in % equivalent primary energy gave us a confusing and unnecessary datapoint for our analysis and given that there were already columns that could be used to address similar points, we dropped this column. The Density column gave an interesting idea for us to address, was there a correlation in a country’s population and energy usage? While this is an interesting question that would be worth looking at in future analysis, we found it to be beyond the scope of the original questions we posed. Furthermore, the column isn’t a total population column, but a density column. Given that countries within our data set range from Singapore, to Malta, to India, we felt that the population density would be something that would crowd the data analysis further without providing value, so this column was also dropped. The column for density had to be drop with the specific columns due to python not recognizing the column with the extra characters were included within the code.



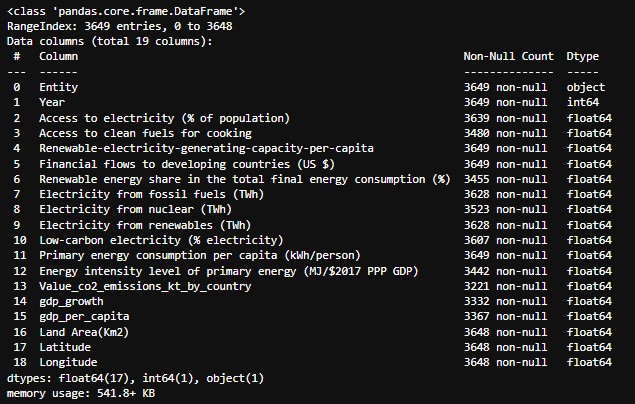


In cleaning the data, we replaced N/A and NULL values with zeros in both the financial flows to developing countries (US$) AND Renewable-electricity-generating-capacity-per-capita columns. We did this as any values that were not given here would only give a more conservative skew on the data.





While this might paint the outcome of the analysis in a more negative light, we felt that with the subject matter, it would make more sense to present data that would look worse and call for more action, rather than an analysis that says there is enough being done, and no call to action. Lastly, we did analysis on a regional basis and to do that, made a new data frame to consolidate the countries by continent for analysis. This feature engineering helped look at a high-level overview of the dataset and how geography can play a factor in the adoption of renewable energy. This left us with a final data set as shown below.

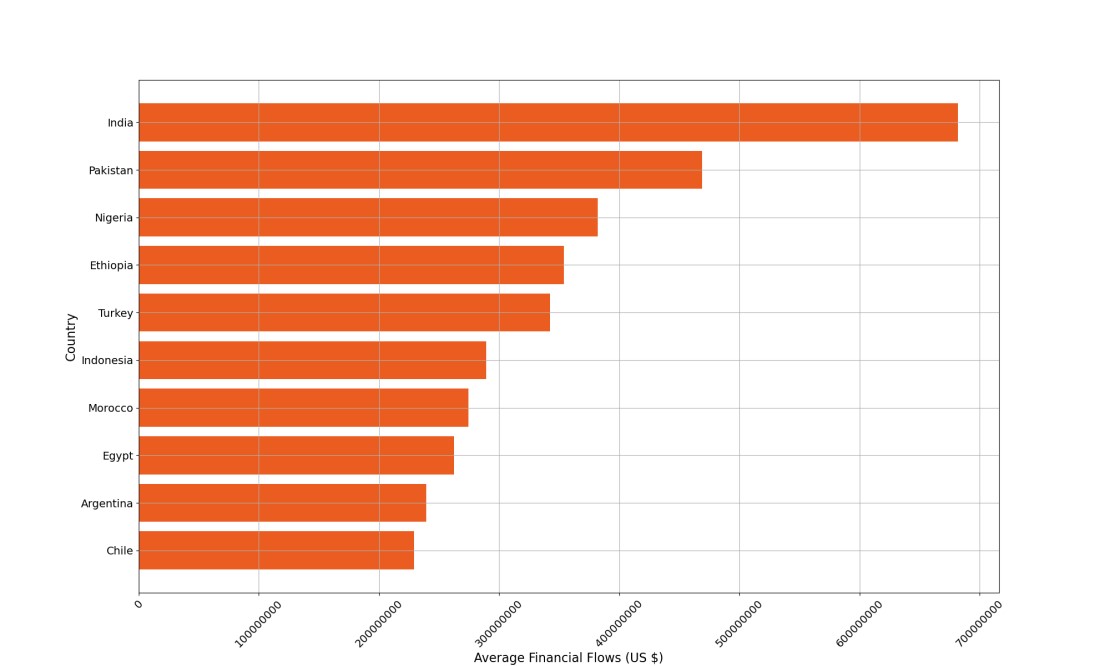


**Question 1: What impact does foreign aid have in developing countries regarding renewables?**

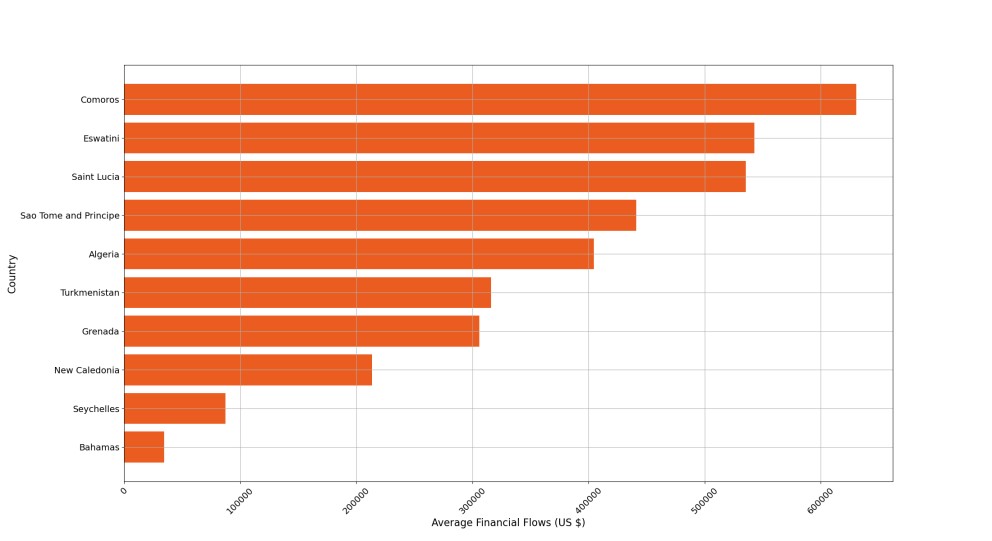
One of the main drivers in the global transition to renewable energies is finance. Developed countries often financially assist with the implementation and upkeep of renewable energy in less developed countries. There are two columns on the twenty-year dataset that our group found interesting to dive into this question further. The financial flows for each country, which shows the amount of funding given to developing countries each year between the year 2000 to 2020 for the purpose of moving towards renewable energy. Also, the renewable energy usage share percentage of a country’s total energy consumption for each year. The goal was to compare these two factors and see if there was any significant correlation between them.

To begin the comparison, the data was drilled down to find both the top 10 countries and bottom 10 countries in both average financial flow and renewable energy usage share percentage over the course of the twenty-year period. Both columns were cleaned of null values and replaced with “0” if there was no information available. Code was run to get the average for both values and which countries were in the top 10 and bottom 10.

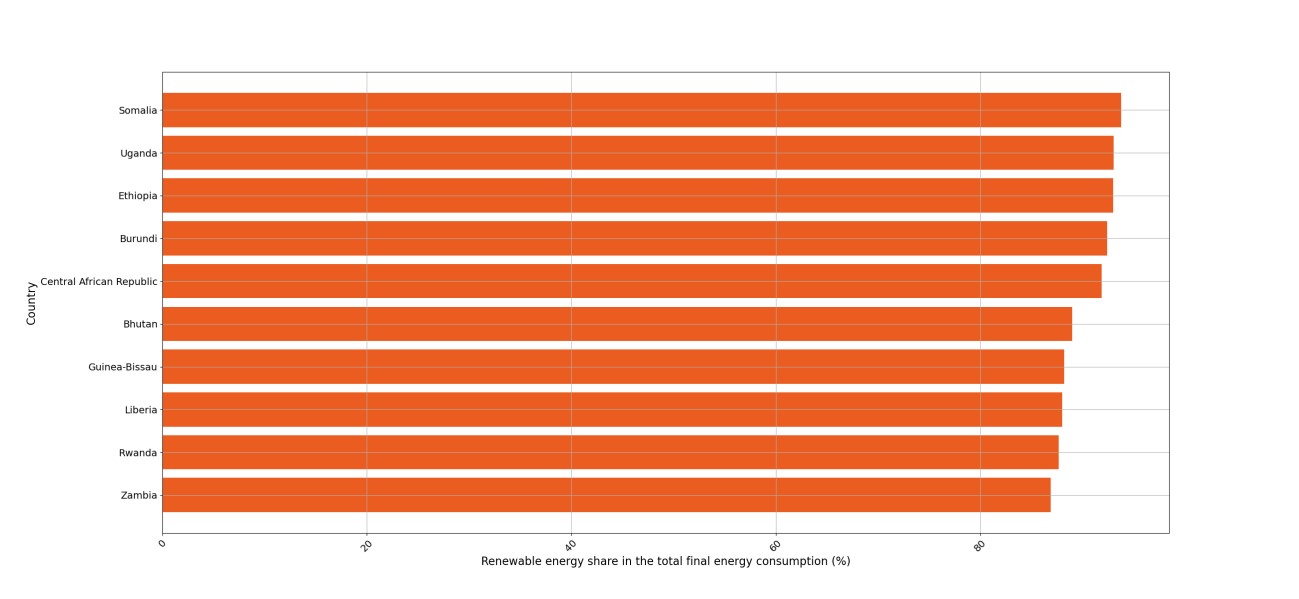
For the financial flows category, developing countries India, Pakistan, Nigeria, Ethiopia, Turkey, Indonesia, Morocco, Egypt, Argentina, and Chile were in the top 10 countries receiving funding from developed countries for the purpose of renewable energy usage on average over the course of the twenty-year period. With each country receiving 100s of millions of dollars on average. Most of the top 10 were made up of countries from both Africa and Asia.



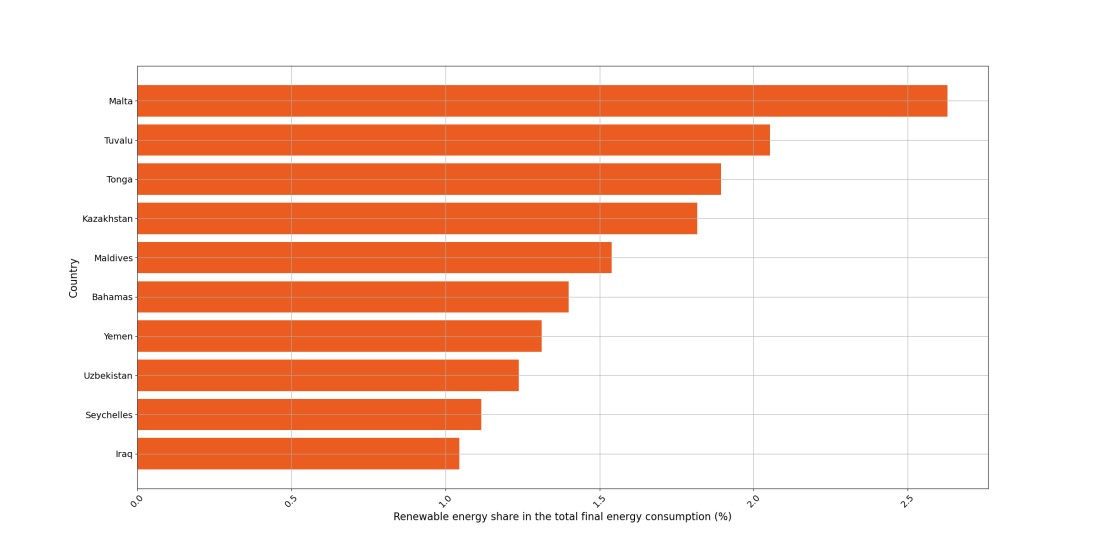
The countries in the bottom 10 that received at least an average of $10,000 in funding during the twenty-year period were Comoros, Eswatini, Saint Lucia, Sao Tome and Principe, Algeria, Turkmenistan, Grenada, New Caledonia, Seychelles, and the Bahamas. All receiving less than $650,000 on average in financial flow. Most of the countries in the bottom 10 were smaller or island countries.



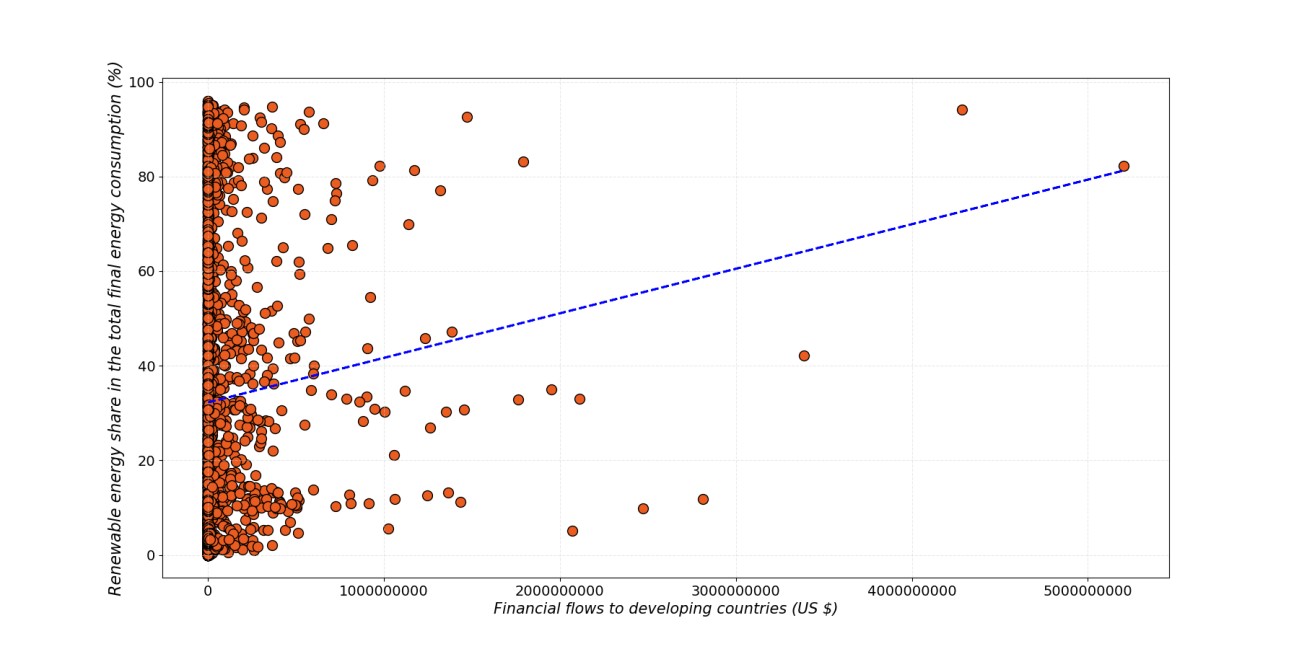
Next, the renewable energy usage percentage category was reviewed. The top ten was made up of countries Somalia, Uganda, Ethiopia, Burundi, Central African Republic, Bhutan, Guinea-Bissau, Liberia, Rwanda, and Zambia. Having an average of at least 80% of energy used being renewable energy during the twenty-year period. Almost all the countries in the top 10 were African countries.



The bottom 10 was made up of countries Malta, Tuvalu, Tonga, Kazakhstan, Maldives, the Bahamas, Yemen, Uzbekistan, Seychelles, and Iraq. Having an average of up to 2.5% of renewable energy being used over the twenty-year period. A decent amount of the countries in the bottom 10 were middle eastern countries.



Next, a scatter plot was used to identify the relationship between these two factors for all the countries in the data set.



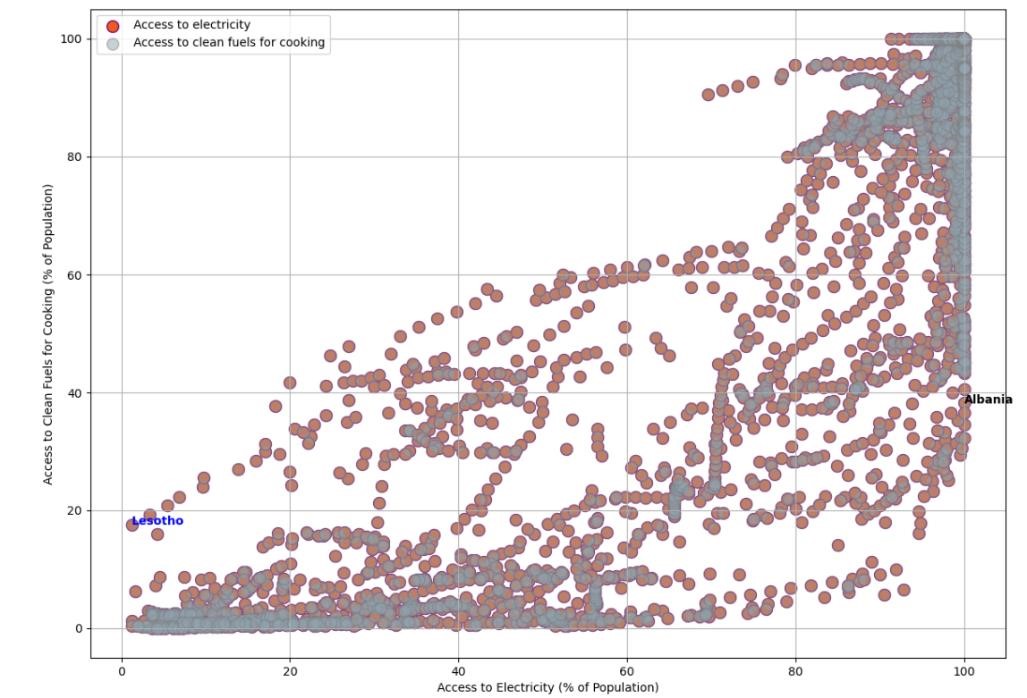
The scatter plot looked at the total amount of funding the countries received over the twenty-year period and their average percentage of renewable energy used compared to overall energy consumption. Most of the countries received very little funding, but many of these countries still had a very high consumption of renewable energy. Also, some countries that received a good amount of funding did not have a very high renewable energy consumption. This gave some insight into the other factors that maybe influencing the pattern of the data that will be explored further in the following questions.

**Question 2: Are there any notable factors that promote the adoption of renewable energy year over year?**

An important driving factor in the adoption in renewable energy happened in 2015, when the UN General Assembly adopted the 2030 Agenda for Sustainable Development and its 17 Sustainable Development Goals known as the (SDGs), which includes a dedicated goal on energy, SDG 7, this is call for action to “ensure access to affordable, reliable, sustainable and modern energy for all".[[3]](#footnote-3) SDG7 is key to the development of many human life center activities such as education, agriculture, business, communications, healthcare and transportation. This graph provides an overview of the different sectors with the most incentives to achieve this goal.

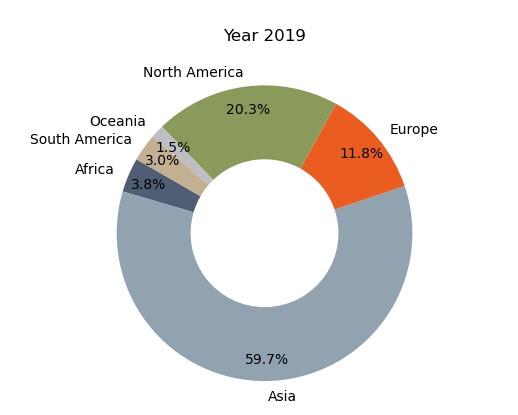
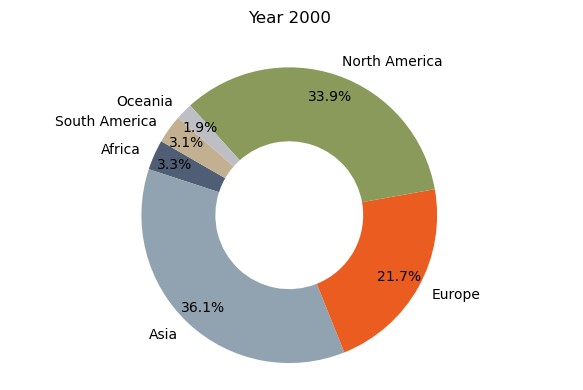


The SDG7 highlights the outside factors the can impact a country’s decision to go with sustainable energy sources or traditional fossil fuel sources. To examine this aspect of this further, we looked at columns such as the Access to Clean Fuels for cooking as a relative gauge of how developed a country is. This could give an indication of a country’s priorities as well; it would make sense for a country to put less focus on sustainable development when their population doesn’t even have access to basic clean cooking fuels. The dataset shows a diversified or mixed progress, for instance, the leader country in the scatter graph, Albania where 100% of its population has access to electricity, has roughly 40% access to renewables. In contrast the dataset shows Lesotho with less than 22% access to both categories. Other countries with high access percentages within the 20-year span had maintained it or even improved their access rates by 2020.

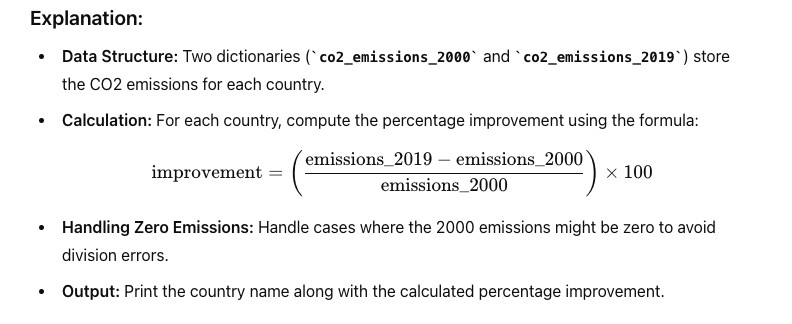


Facts around this data suggest that somecountries with consistent improved access rates may have implemented effective policies, had better infrastructure, and had success on their technology implementation and awareness campaigns efforts. In addition, the regional growth inconsistency calls for a significant need for improvement for the countries that still lagging. Efforts to address these disproportions are crucial for achieving global access towards clean energy goals.

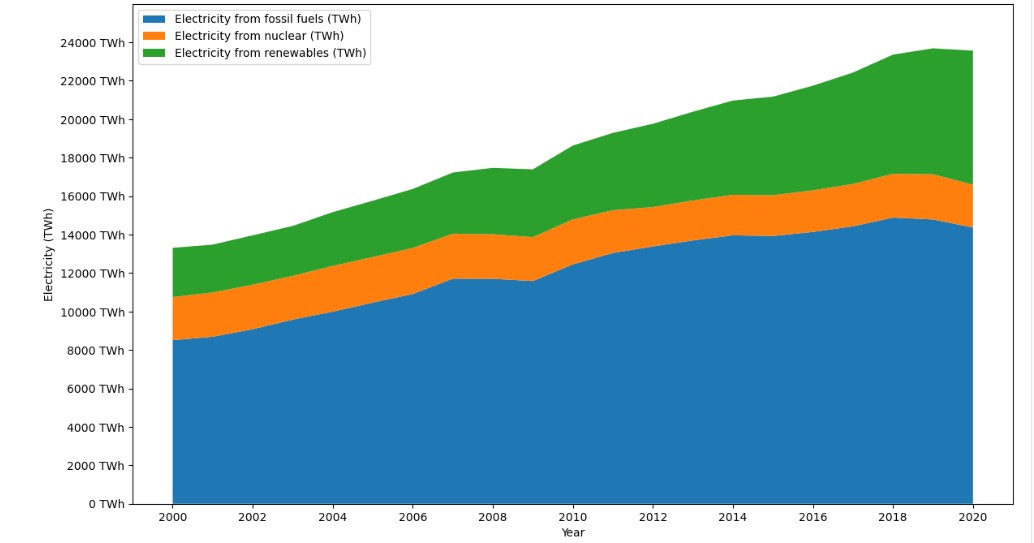
**CO2 Emissions Distribution**



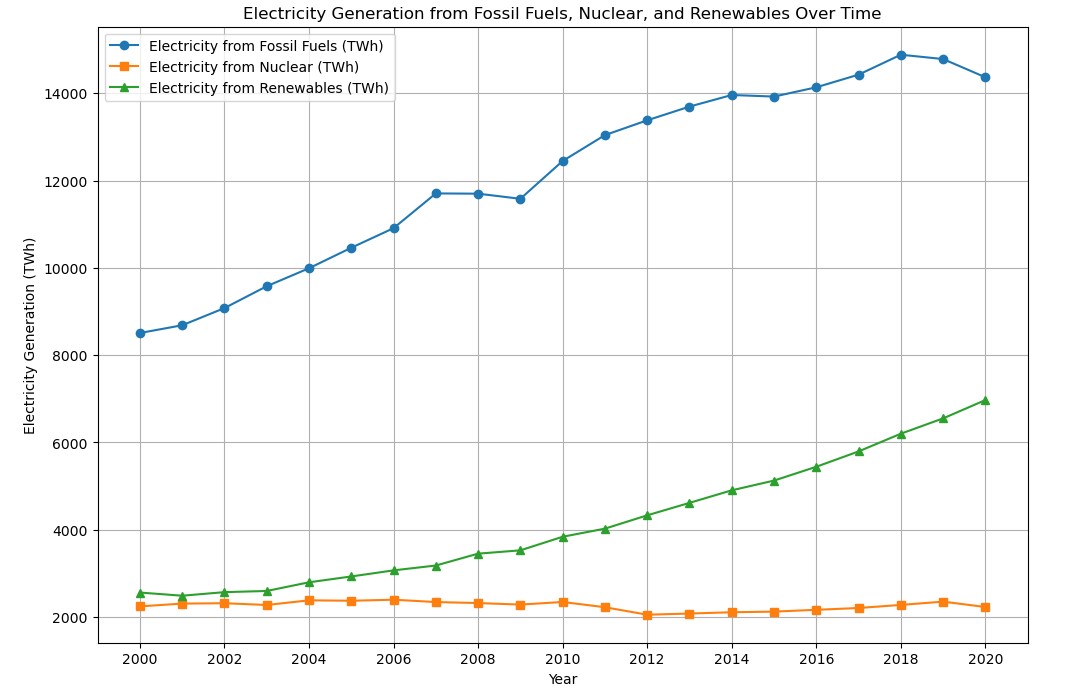
The doughnut charts show a comparison per continent. We selected the year 2000 as starting point and then compared against 2019, interesting enough our dataset did not provided data for the 2020 year (unfortunately). As a result, you can see 19.89% increase in the 19-year span. Also, it is noticeable in contrast to a significant increase in CO2 emissions for continents like Asia and Africa.



**Electricity Generated by Source [[4]](#footnote-4)**



Under the electricity from renewables our dataset did not make a distinction on the type of renewables, the mixed renewables include solar, wind, hydropower and bioenergy among others. As you can see there is a progressive increase across the board, and is encouraging to see renewables gaining ground, but there are still ways to go.

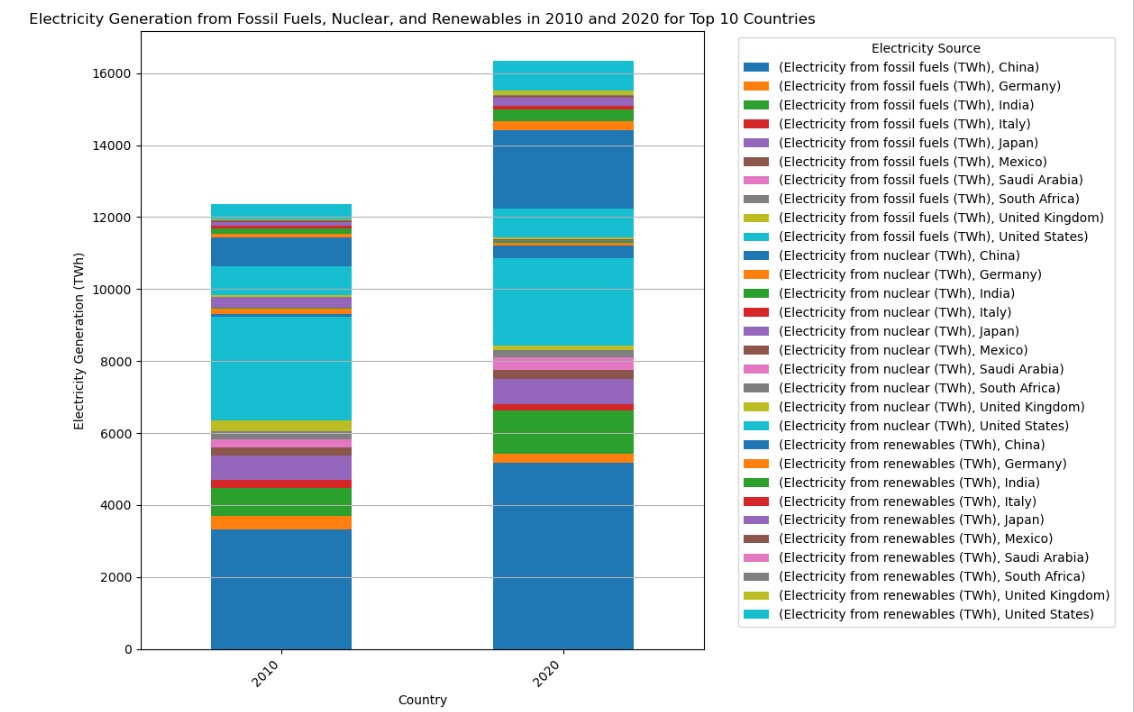


There are two decrease points in our dataset:

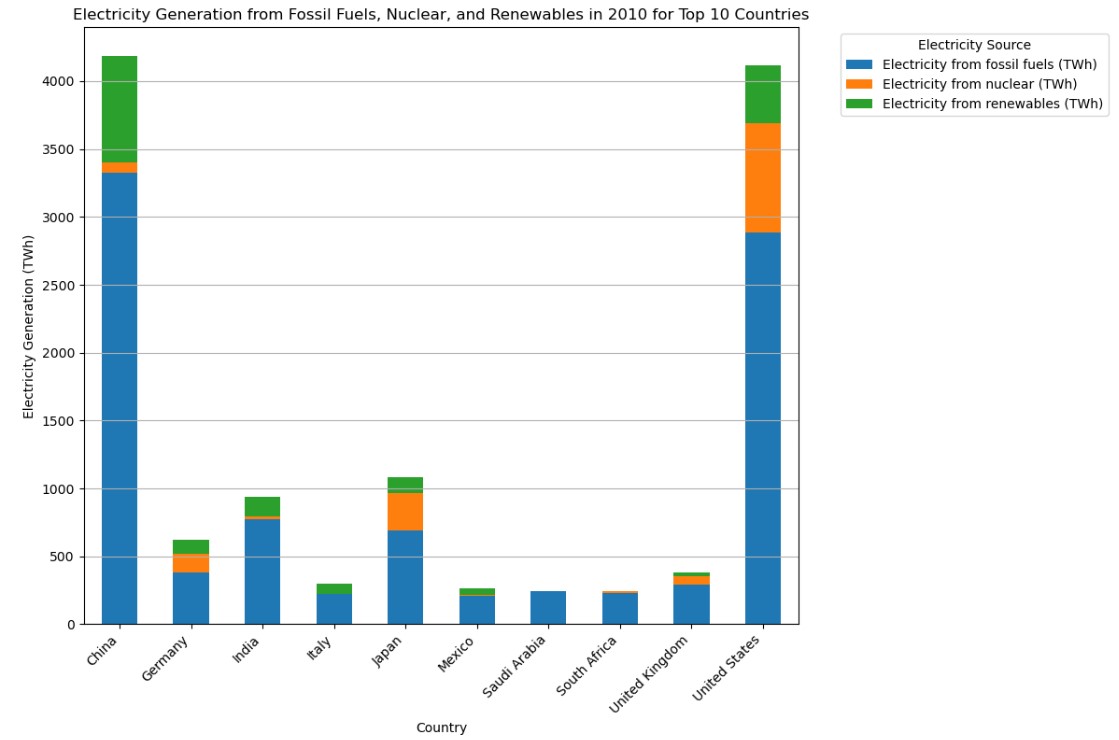
1. In 2009 There was a significant decline in the global production and consumption of oil.

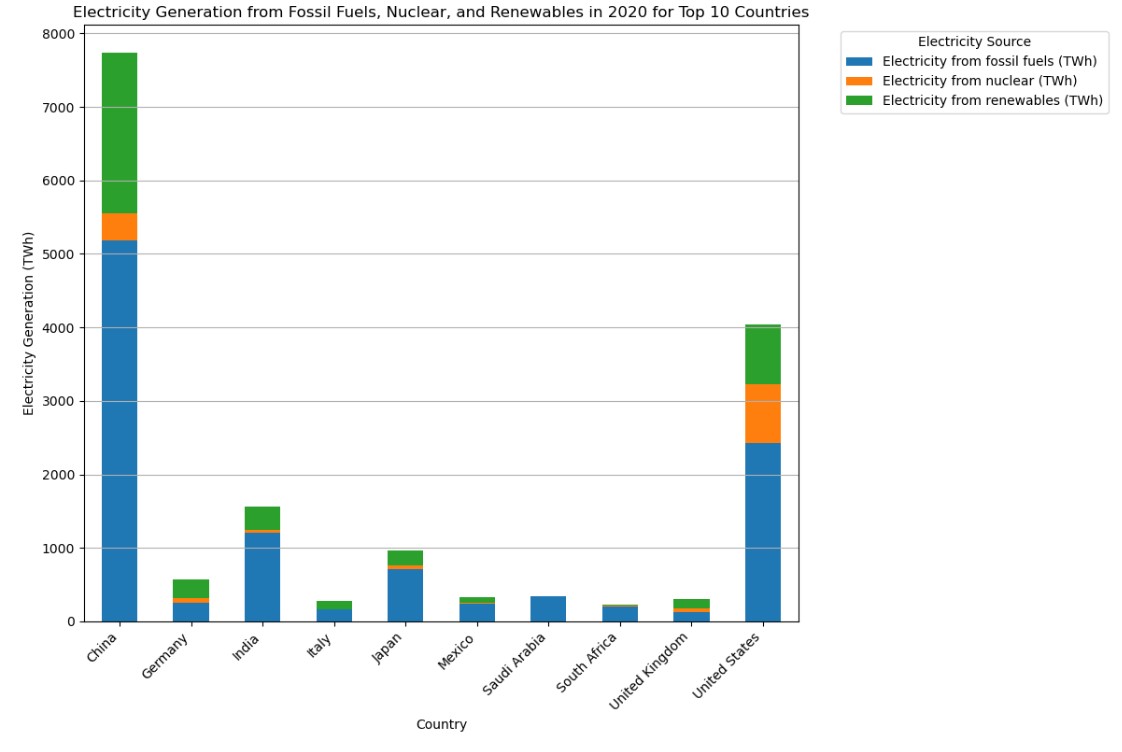
which was impacted by a global recession. Also, you may recall the famous effects of the 2008–2010 automotive industry (bail out) US crisis.

1. In 2020 Fossil fuels and nuclear electricity generation experience a direct impact due to the coronavirus pandemic. Lockdowns worldwide reduced the demand for gas needed for transportation and other daily activities. On the upside however the generation of renewables showed an increase in that same year.



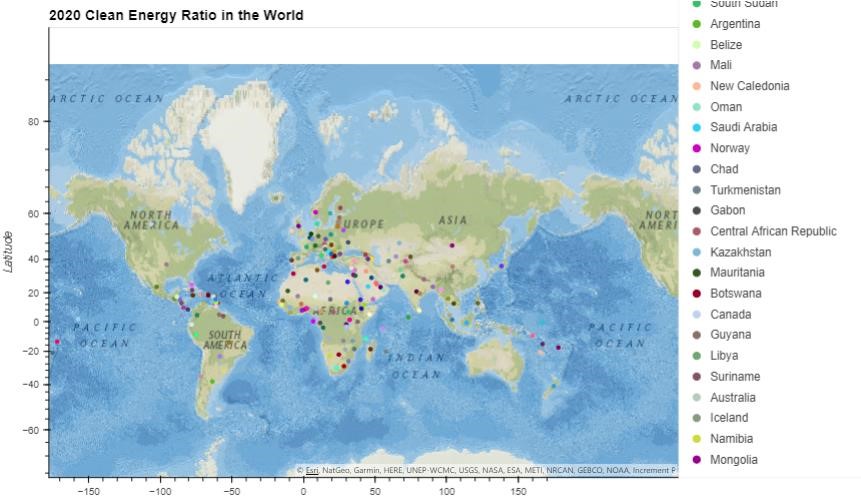
The emergence of solar energy as a viable and competitive energy source has played a significant role in the transition towards a more sustainable and renewable energy future. It has diversified the energy mix, reduced greenhouse gas emissions, and provided opportunities for individuals, businesses, and communities to generate their own clean energy. Below is a detailed bar graph with the top 10 countries’ higher energy producers and their participation. Overall, the data comparison for two decades highlights both progress and challenges in ensuring widespread access to clean fuels for cooking globally. Continued efforts and investments are needed to address the remaining gaps and ensure that all populations have access to clean and sustainable cooking options. indicating the need for further infrastructure development and investment.





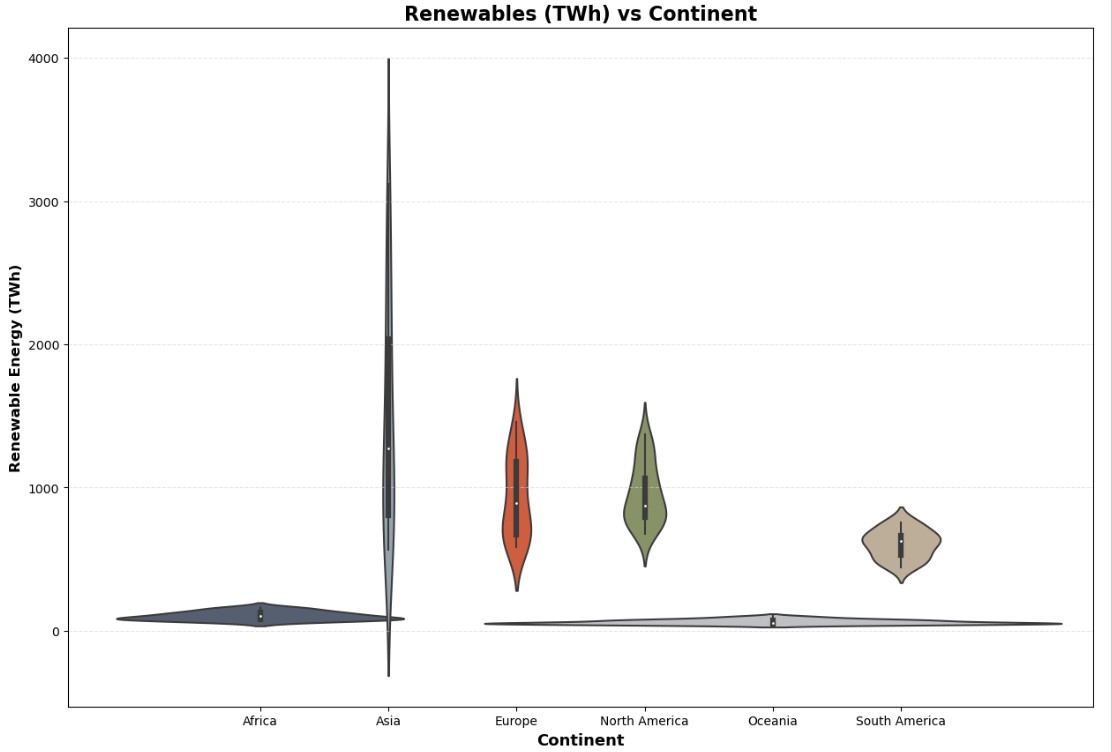
**Question 3: What conclusions can be drawn by examining a country’s renewable usage by geographical location?**

Our first step with examining this question was to break the dataset into a regional basis as described in our data engineering. Below is a map showing the clean energy ratio by country for every country within the data set. The primary intent of the color points on the map was to show an increase in renewable energy ratio from 2000 to 2020 by change in color and points size. When analyzing this we determined that the map was not conclusive enough due to the visual skew of the map given the number of countries that are close to the equator and the size of the datapoints being relative to the clean energy ratio did not provide a valuable enough insight to draw a meaningful conclusion from the visual. Alternatively, this shows countries with installed renewable energy and higher concentration of renewable energy closer to Equator. This might be because of higher installed solar capacity.



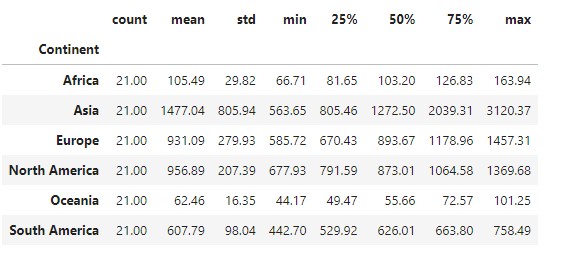
To better understand the data the countries were grouped by continent, with the feature engineered we were able to better understand this data on a regional basis and give visuals that were more impactful and valuable in drawing conclusions regarding the dataset.

Looking at renewable energy by continent, Asia has the highest installed renewable energy. This is skew by India and China. Violin Chart compares median of each region, with Asia having highest median followed by Europe and North America. This immediate conclusion shows the impact of grouping by region and how looking at aspects of this data set at an overall country level can very highly skew certain data points and conclusions.

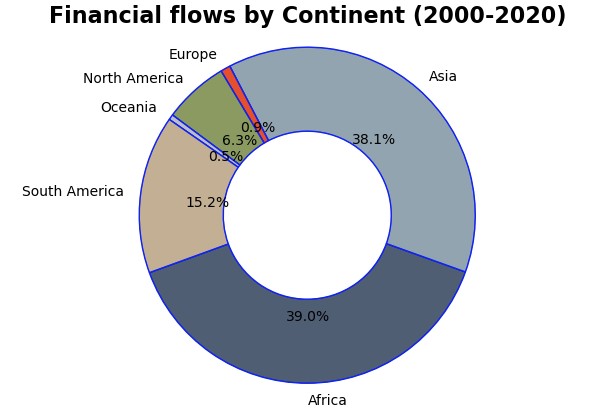


As can be seen by statistical analysis Asia has a mean of 1500Twh followed by Europe & North America of 900Twh and Oceania the lowest 60TWh installed renewable energy capacity.

Given general knowledge of the regions and the countries within them this would make sense. While further considering that Africa has a similar distribution to Oceania, while having a significantly higher population, land mass, and number of individual countries. Giving us a high indication that there is significant room for improvement here. Whereas the distribution of Europe and North America are in line with one another, a more expected result.



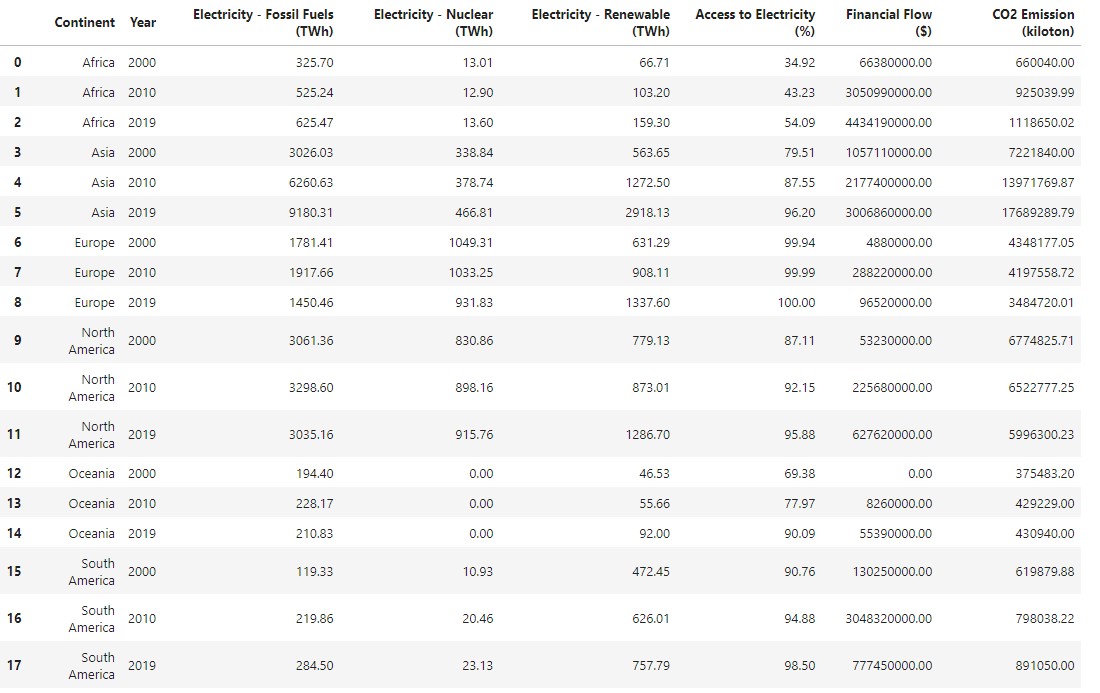
The below donut chart shows that Asia and Africa received about 80% of funding. India and Pakistan received maximum funding in Asia, while Ethiopia and Nigeria received more funding in Africa. As a result, Asia & Africa’s installed Renewable energy capacity increased by about 5 and 3 times respectively. This contradicts the micro level analysis we discussed earlier. By looking at it from higher level /continent level it can concluded that the financial funding has helped developing countries to move towards renewable energy.



Individually we have discussed the dataset in terms of percentages and statistical data, table below shows overall energy and financial flows by continent absolute data comparing the year

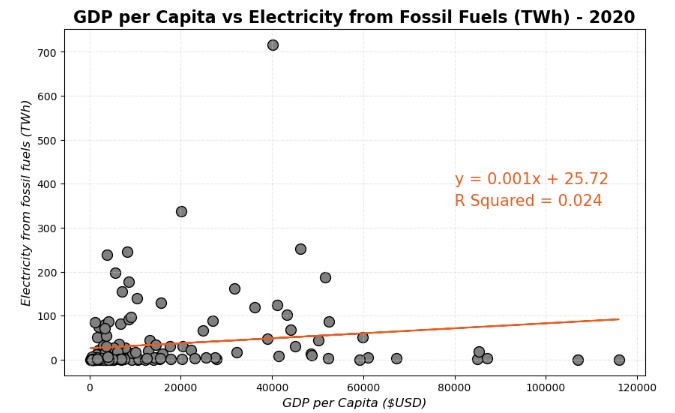
2000/2010/2019 by continent. It can be concluded from table -

1. Nuclear energy has not changed for most part because of risk associated with the technology
2. By focusing on Africa and Asia if Renewable energy was not considered this would have resulted in a 30% increase in C02 emission in Asia and a 25% in Africa assuming linear relationship between CO2 emission and installed electricity by fossil energy
3. Point worth noting here is access to electricity (col 5) shows a 50% of Africa still do not have access to electricity some of this can possibly be offset by solar or wind

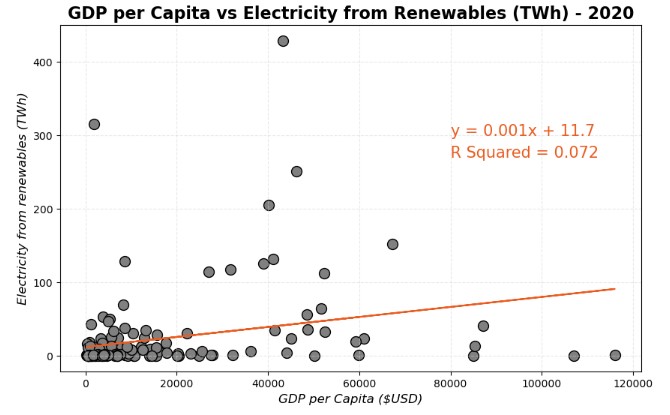


# Regression

The regression that we went with originally was a comparison of the financial flows to a country and their total share of energy from renewables. After reviewing the correlation between these though, we found that this would be an ineffective regression. The correlation on these was negative. Using this insight, we pivoted to a regression with a higher correlation to attempt to get an impactful result from the regression. Pivoting from our original idea we chose to do two related regressions with a higher correlation, GDP per Capita vs Electricity from fossil fuels (TWh) and GDP per Capita vs Electricity from Renewables (TWh). Both regressions were done only with the 2020 data from our dataset to use the most current data to regress.



They both have low R-Values but can still over valuable insights, having a large left skew we can conclude that the number of countries with lower relative energy consumption and GDP could be further analyzed within their own data set if there was a criterion established to further categorize the dataset. There is also not enough of a correlation here to make this a predictive model which, while this was a goal of this project, offers another valuable insight. That the outside factors analyzed in question two play a very significant role in how and if a country adopts renewable energy.



# Bias and Limitations

When considering our dataset are several bias and limitations that are found within that need to be considered when drawing conclusions. It is limited by the fact that there are a high number of socio-economic factors that would impact datapoints across this analysis that aren’t necessarily quantifiable with this type of dataset. A countries individual politics and polices would play into this point as well. The dataset also did not include all countries of the world. 148 out of 195 were included so the sample size is still more than enough to draw conclusions from, but it must be stated that this is a limitation of the data. There were also several columns withing the data set that reference ‘developed’ or ‘developing’ nations but the data set nor the information on it within Kaggle defined what each of these tags meant. This leaves room for bias within the dataset on how these are defined. The dataset also does not quantify factors such as how funding for renewables is used to develop an infrastructure within a country. There is a higher level of infrastructure needed to utilize sustainable technologies in most cases, so if a

country does not have this, how effectively are they able to adopt sustainable technologies?

Finally, a breakdown of the types of renewable energy is missing. This will help to show the concentration of wind, solar, hydro and other forms of energy and would be useful in determining if certain regions are able to utilize different types of sustainable energy in a more effective way.

# Conclusion

Reviewing our analysis, we find that there has been very mixed progress overall in the last 20 years in the development of sustainable technologies. While renewables usage has gone up, so has total energy consumption, were believe that a good next step for a study in this field would be to analyze the ratio of sustainable energies used and how that ratio has changed over time. Because the cost of implementing renewable energies is higher than the cost of implementing traditional energy sources, there needs to be a higher focus on decreasing that barrier to entry for lower GDP countries so that funding can be more effective in impacting the percentage of renewables that are utilized by a country and financial support can have a more direct impact. Though the effectiveness of funding is not high, further work needs to be done in the analysis to show where funding is best used and which countries that it sent to in a county or region that don’t have a good effective rate of utilization.

# Bibliography

1. Tanwar, Ansh. “Global Data on Sustainable Energy.”, Kaggle, www.kaggle.com/datasets/anshtanwar/global-data-on-sustainable-energy/data. Accessed 6 June 2024.
2. “42 U.S. Code § 18761 - Definitions.” Legal Information Institute, Cornell Law School, www.law.cornell.edu/uscode/text/42/18761#a\_1. Accessed 6 June 2024.
3. Lacey, Stephen. “Chart: Renewables Generated a Record 30% of Global Electricity in 2023.” Canary Media, 12 Jan. 2024, www.canarymedia.com/articles/clean-energy/chart-renewables-generated-arecord-30-of-global-electricity-in-2023. Accessed 6 June 2024.
4. Chakravarty, Shoibal, et al. “Scatter Plot of per Capita CO2 Emissions as a Function of GDP per Capita for the Sampled Countries.” ResearchGate, www.researchgate.net/figure/Scatter-plot-of-percapita-CO2-emissions-as-a-function-of-GDP-per-capita-for-the-sampled\_fig5\_338650331. Accessed 6 June 2024.
5. Ritchie, Hannah, and Max Roser. “Renewable Energy.” Our World in Data, 2024, ourworldindata.org/renewable-energy. Accessed 6 June 2024.
6. OpenAI. "ChatGPT Response to Question about MLA Citation." Accessed 6 June 2024, chat.openai.com.
7. Sustainable Development Goals [https://www.un.org/sustainabledevelopment/energy/#:~:text=Goal%207%20is%20about%20ensuring ,targets%20%E2%80%93%20but%20not%20fast%20enough.](https://www.un.org/sustainabledevelopment/energy/#:~:text=Goal%207%20is%20about%20ensuring,targets%20%E2%80%93%20but%20not%20fast%20enough)

8.Recession drove 2009 energy consumption lower. https://www.bp.com/en/global/corporate/news-andinsights/press-releases/recession-drove-2009-energy-consumption-lower.html

9.Effects of the 2008–2010 automotive industry crisis on the United States.

https://en.wikipedia.org/wiki/Effects\_of\_the\_2008%E2%80%932010\_automotive\_industry\_crisis\_on\_the \_United\_States

10. Global Energy Review 2021 https://www.iea.org/reports/global-energy-review-2021/renewables

Assessing the effects of economic recoveries on global energy demand and CO2 emissions in 2021.

1. Tanwar, Ansh. “Global Data on Sustainable Energy.”, Kaggle, [www.kaggle.com/datasets/anshtanwar/global-data-on-sustainableenergy/data. Accessed 6 June 2024.](http://www.kaggle.com/datasets/anshtanwar/global-data-on-sustainable-energy/data.%20Accessed%206%20June%202024)

   [↑](#footnote-ref-1)
2. Sustainable Development Goals https://www.un.org/sustainabledevelopment/energy/#:~:text=Goal%207%20is%20about%20ensuring,targets%20%E2%80%93%20bu t%20not%20fast%20enough.

   [↑](#footnote-ref-2)
3. Global Energy Review 2021 https://www.iea.org/reports/global-energy-review-2021/renewables Assessing the effects of economic recoveries on global energy demand and CO2 emissions in 2021. [↑](#footnote-ref-3)
4. This graph shows the progression of electricity generated by the 3 types of energy over 20 years in terawatt hours (unit of energy representing trillion-watt hours). [↑](#footnote-ref-4)