Analysis and Prediction of Sustainability and Utilisation of Energy Sources

Project Report for *Indian Institure of Technology, Bombay* - DS203: Programming for Data Science (2021)

Tanmay Dokania
Dept. of Electrical Engineering
Indian Institute of Technology, Bombay
Mumbai, Maharashtra
200070083@iitb.ac.in

Navneet
Dept. of Electrical Engineering
Indian Institute of Technology, Bombay
Mumbai, Maharashtra
200070048@iitb.ac.in

Pranava Singhal
Dept. of Electrical Engineering
Indian Institute of Technology, Bombay
Mumbai, Maharashtra
200070057@iitb.ac.in

***Abstract*—The quality of human life, the efficiency of our civilization, and the power behind our technology all come from energy. Energy usage varies widely across the world and shows surprising relations with the quality of life and economic well-being in these countries. We analyse the usage of a wide variety of energy sources, the trends in carbon emissions and its relationship to global climate policy, the disparity between rich and poor countries when it comes to energy access and lastly we draw projections for carbon levels in the near future. We analyse some carefully chosen countries in more detail especially with regard to the evolution of renewable usage in recent years. The energy data is sourced from British Petroleum [1] while the data for GDP and HDI development metrics are sourced from UNDP [2].**

I. Introduction

Energy demand has been growing rapidly all over the world to meet the needs of an exponentially rising population and to keep up with the needs of our technological progress. While most energy has been traditionally sourced from non-renewable sources like oil, coal and natural gas, other renewable energy sources are becoming increasingly important. These renewable energy sources are vital since the carbon footprint of conventional energy sources is unsustainable and their supply is limited. The availability of energy is strongly tied with the progress of a nation since power is needed to run everything from industries to households.

With a large number of energy sources and large and unpredictable variations in their usage, analysis of global energy usage is challenging. Every country offers unique circumstances that make the usage of certain sources more economical and geographically-suited as compared to others. This analysis however is essential because it enables policy-makers to make informed decisions about global climate and energy policy and ensure a sustainable future. The tools of data science help us understand the current trends in energy usage and carbon emissions as well as help us make future predictions that are essential for guiding global policy.

The existing research motivated us to explore relationships between energy usage and well-being of individuals. This led us to analyse correlations between key development metrics like Gross Domestic Product (GDP - per capita) and Human Development Index (HDI). Economics does play a major role in influencing how a country uses its energy and we analysed these patterns by combining energy data with development data to draw interesting observations. We also looked at the impact of some key global events on CO_2 levels and energy usage.

Our data primarily comes from two sources. The energy data has been taken from bp Statistic's review of World Energy [1] while data on GDP and HDI (development metrics) has been taken from United Nations Development Programme, Human Development Reports [2]. We have linked together these data sources.

In this project we have primarily done the following things:

- We have performed extensive data cleaning and filtering to reduce a large dataset to a compact and easily usable form
- We have analysed the correlation between development metrics and energy usage to understand how economic well-being impacts people's consumption of energy.
- We have analysed the distribution of energy coming from different sources and analysed trends in this distribution over time for the world as well as some carefully chosen countries. A country's energy usage reflects several things about its policies, initiatives and economics.
- We have tried to gauge the impact of past policy interventions and key global events on CO_2 and energy consumption levels.
- We have used machine learning techniques to draw projections for future CO_2 levels and energy usage.

Our primary aim was to identify countries, policies and economic circumstances that enable a sustainable and happy future for people through conservative energy usage and limited carbon emissions. However, it is challenging to give policy recommendations based on this analysis alone since every country's needs are different. Our study highlights the wide variation in energy usage across the world and hopefully makes us realise that we have a long way to go in reaching sustainable levels globally.

II. Prior Work

With growing concerns about climate change and our rising demands for energy a lot of work has been done in this field. British Petroleum, through bp Statistics' Review of World Energy [3] has carefully collected and thoroughly analysed the key trends in energy usage across the world, CO_2 emission levels and key minerals, to name a few. The Special Report on Emission Scenarios (SRES) by the Intergovernmental Panel on Climate Change (IPCC) [4] has created projections for future CO_2 levels based on considerations of possible global policies and is a fairly complex model.

III. Dataset and Methodology

The project deals with describing the dependence of different energy sources and their growth from the year 1965 to 2020. We primarily used data concerned with energy usage, CO_2 levels, GDP and HDI for all countries in the world.

A. Datasets

World Energy Statistics: We have taken this data from the Statistical Review of World Energy 2021 - BP(British Petroleum) [1]. This data shows energy consumption/production of energy from various sources of several countries from the year 1965 to 2020. We used this data to draw conclusions on the growth of renewable energy dependence in different countries, CO_2 emissions, etc. Some of the sources of energy that are considered for analysis are : "Coal", "Oil", "Natural Gas", "Solar", "Wind", "Nuclear", "Hydro Power", "Geothermal", "Biomass", "Biofuels". CO_2 emissions are also considered to analyze their time evolution in different countries. While this dataset has information about most countries, data for large parts of Africa and a few other countries is missing. We have dropped those countries which had negligible or no data from our analysis. This is regrettable as it makes our analysis incomplete and biased.

GDP per Capita and Human Development Index (HDI): This data is taken from the United Nations Development Programme (UNDP) [2] which publishes key data about global development in its annual Human Development Reports. We have borrowed data for Gross Domestic Product per capita, which is a measure of the cost of all final goods and services produced in a country. It is a standard metric for the economic well-being of a country. While GDP does tell us about the market conditions and economic strength of a nation it is not an adequate measure of human happiness. Economics is not the only factor driving human satisfaction and thus, we have also borrowed data for the Human Development Index (HDI) which is a more comprehensive measure of well-being. It takes into account the quality of education, health and work along with several other human factors.

B. Data Pre-processing

First, we started by cleaning our data and reducing it to relevant information. To do this we removed those countries from our datasets that don't have sufficient data to draw conclusions. Every energy source has multiple data columns such as production and consumption in various units and some columns have mineral availability as well. We only kept the columns of relevance for our analysis and dropped the rest. In further analysis we also dropped columns and rows with excessive NaN values since the presence of NaNs make data interpretation and visualisation harder downstream.

The columns of the filtered dataset were first transformed into the same units for appropriate comparison and then converted into per capita values for each country. After per capita conversion we merged the GDP and HDI data with our main dataset only for the countries present on the main dataset. GDP and HDI data were borrowed from a different source and thus, several conflicts between data sets had to be resolved manually (such as different naming styles for the same country). Further, we also transformed GDP and HDI data to include a single column for years in place of the original data which had a separate column for each year. 'Country-Year' pairs form a natural index for all our datasets and thus make it easy to merge the three dataframes in python.

IV. Data Visualisation

A. Variable Correlation

We are faced with a large amount of data and thus, it is essential that we ask ourselves how we can make sense of this data. The first task was to identify whether per-capita metrics or total metrics are relevant for our analysis.

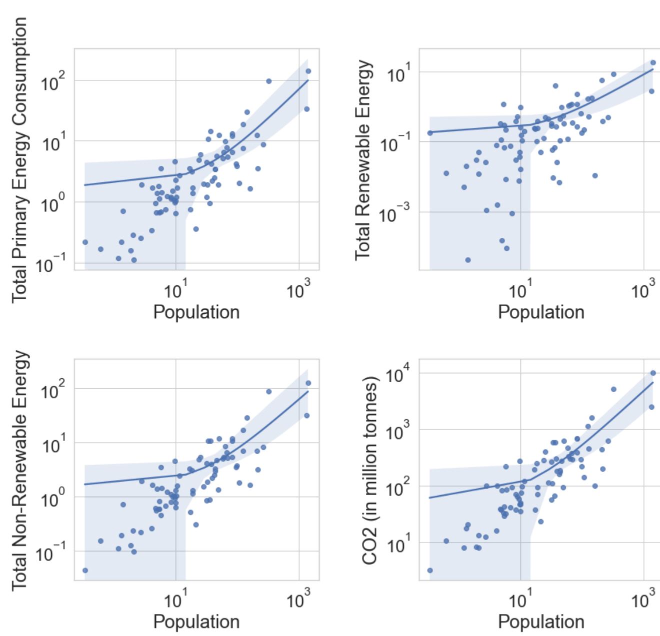

Fig. 1. Correlation between total metrics and population

We justify the usage of per capita data for analysis for which we plot regression plots of total metrics and per-capita metrics against the population.

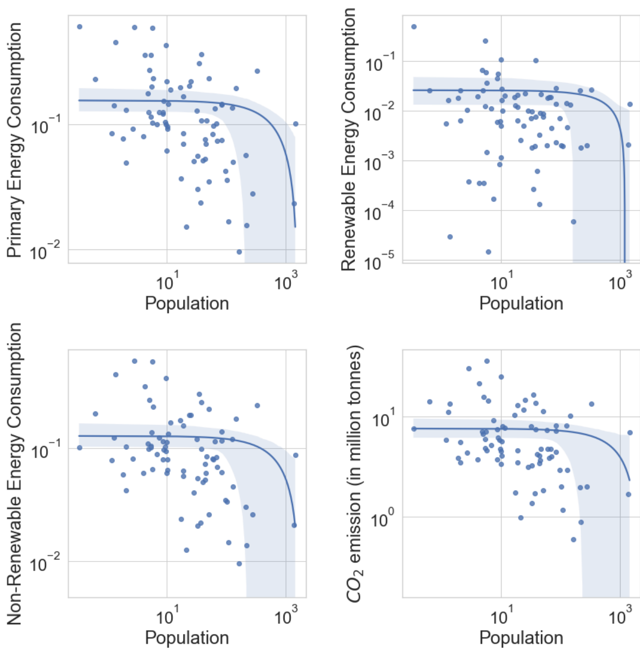

Fig. 2. Correlation between per capita metrics and population

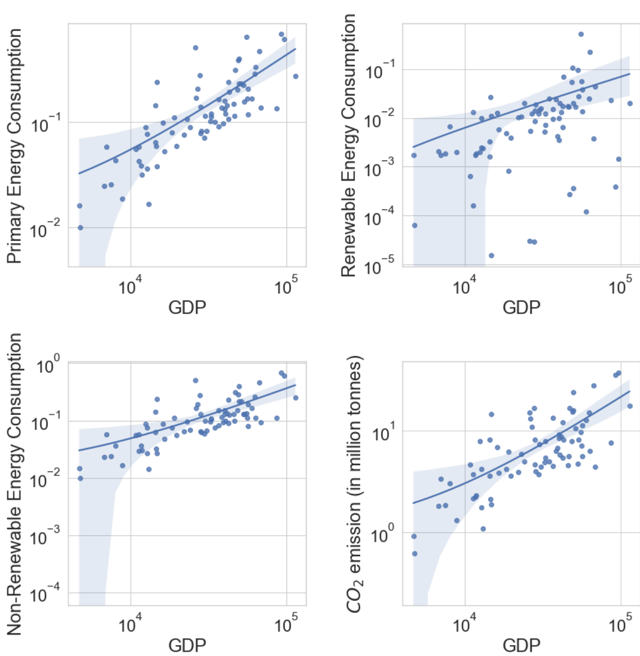

Fig. 3. Correlation between Per-capita Metrics and GDP

We observe that total metrics show an increasing trend with population while the per capita plot gives a nearly uncorrelated plot with population and hence a better inference of a country's conditions. For larger populations, however, a dip in per-capita measurements of energy usage and emission suggests that the average person in high population countries has a smaller carbon footprint. This is further explained by the negative correlation between GDP and population for high populations, confirming the idea that less developed countries often have larger populations.

The regression plots of per-capita metrics vs development indices like GDP and HDI, convey the fact that more developed countries (as measured by these metrics) enjoy a higher standard of living and thus, individuals consume more energy and also generate more CO_2 on average.

We also analysed the correlation of different parameters by plotting a correlation heatmap. We make the following key observations:

- All pairs of variables in the following list are positively correlated - Primary energy consumption, non-renewable energy consumption, CO_2 emissions and GDP.
- The following pairs of variables are negatively correlated - Population with GDP, Population with HDI.

Since most energy use around the world remains non-renewable, it is strongly correlated with total energy consumption. Moreover, wealthier nations consume more energy per person and emit more CO_2. In order to generate the correlation heatmap we have taken care to shift the colour map with its lightest shade centred at zero. This ensures that positive and negative values are visually symmetric.

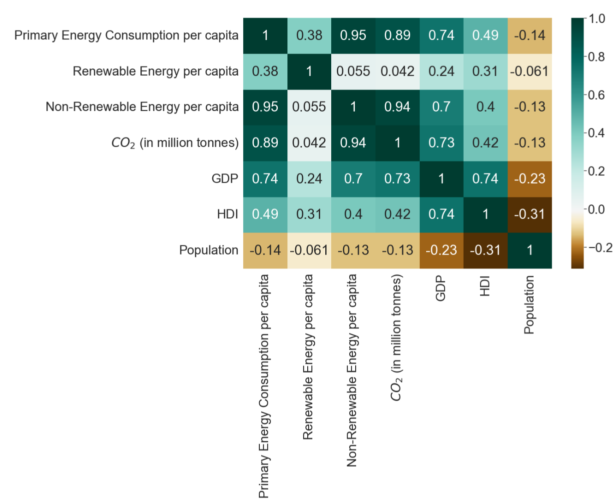

Fig. 4. Correlation Heatmap

B. Variable Distribution

The frequency distribution of the development metrics and per capita energy consumption conveys that few countries enjoy a high per-capita GDP and high per-capita energy consumption, both of which are metrics for the well-being of individuals. However, the graph for HDI is right skewed which is indicative of the fact that despite low economic output and energy usage many countries enjoy a good quality of life as measured by HDI. HDI is considered a more holistic metric since it takes into account the quality of health, education and several other human factors.

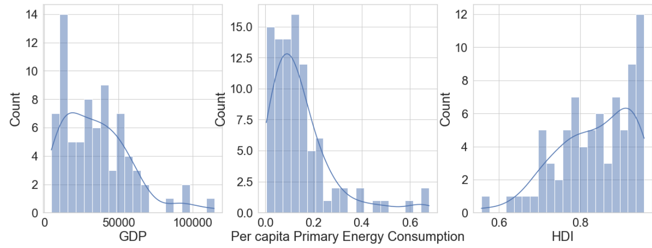

Fig. 5. Distribution of Well-being in the World in 2019

Another key variable of interest is the fraction of energy that is obtained in a renewable manner in a country. Visualising this on the world map tells us that most countries get a small fraction of their energy from renewable sources. only a few countries occupy the 30% to 85% band including the Nordic countries, parts of South America, Canada, New Zealand, and Spain to name a few.

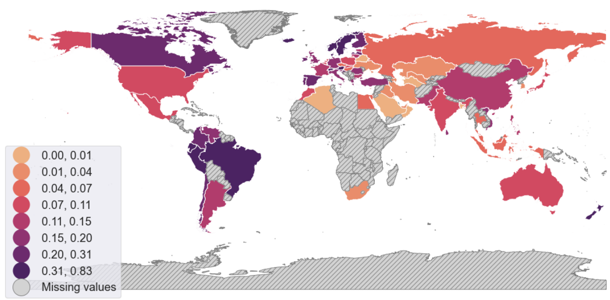

Fig. 6. Distribution of fraction of Renewable Energy

V. Experiments and Results

A. Identifying countries at the extremes of environmental footprint

To analyse the per capita utilisation of different types of energy sources and trend of CO_2 emissions we identified top leading and trailing countries in various categories as follows: (Top and bottom 5 countries by)

- Total Primary Energy Consumption per capita (renewable and non-renewable)
- Renewable Energy Consumption per capita
- Non-renewable Energy Consumption per capita
- Fraction of Energy Obtained from Renewable Sources
- CO_2 emission per capita

Two key results of interest here are the countries with maximum fraction of energy obtained from renewable sources and countries with maximum and minimum per capita CO_2 emission. We will analyse each of these cases in more detail.

The countries that obtain most of their energy from renewable sources are

1) Iceland *(more than 80% of energy from hydroelectricity and geothermal energy)*
2) Norway *(more than 65% energy from hydroelectricity)*
3) Sweden
4) Brazil
5) New Zealand

The countries that feature at the bottom of the list are mostly oil-producing countries. The fraction of energy consumed by them in a renewable manner is so tiny that we had to plot the data on a log scale to visualise it.

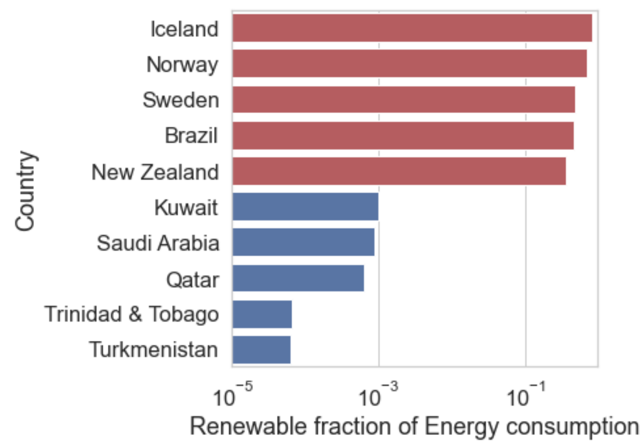

Fig. 7. Top Bottom 5 Countries by Preference for Renewables in 2020

The countries that emit the most CO_2 per capita are

1) Singapore
2) Qatar
3) United Arab Emirates
4) Kuwait
5) Saudi Arabia

As we will see many of these countries are heavily reliant on non-renewable energy sources, which explains the high CO_2 emissions to a large extent.

The countries that emit the least CO_2 per capita are

1) Peru
2) Philippines
3) Sri Lanka
4) Pakistan
5) Bangladesh

We can see that all of these are among low-income countries due to which their per-capita energy consumption and CO_2 emission are both significantly low. These observations confirm that the correlations observed earlier hold good for many countries.

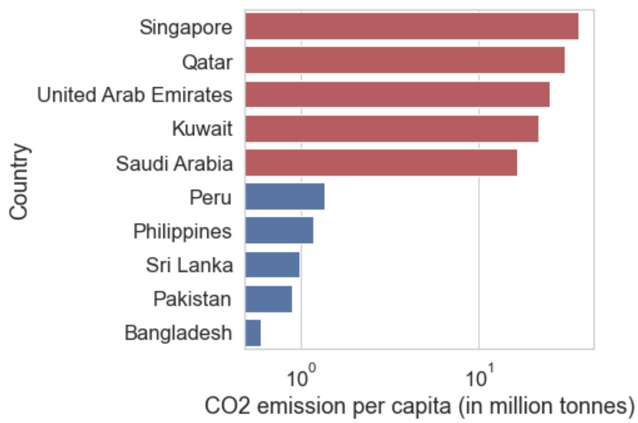

Fig. 8. Top Bottom 5 CO_2 emitters in 2020

B. Analysing countries based on GDP

Above analyses indicate that economic factors and energy consumption are strongly tied. Let us understand what relationships exist between the GDP of a country and its energy usage. We will keep the global per-capita metrics as reference while analysing all countries.

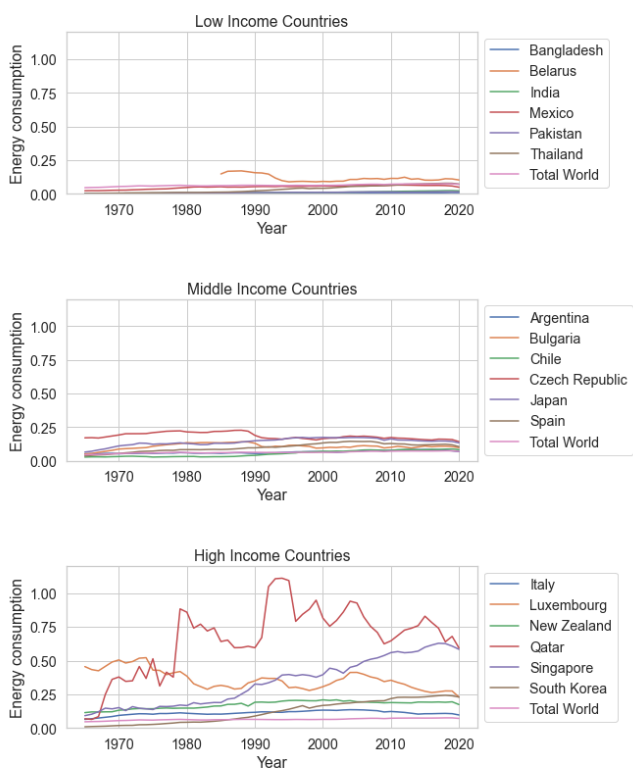

Fig. 9. Per-capita Energy Usage compared to the World Energy is in Exajoules

First we divide countries into three quantiles based on GDP. For simplicity we will pick the top 3 and bottom 3 countries by GDP within each group so as to capture the broad trend. While we may miss out particular countries in this analysis, the goal of this analysis is to look at groups together rather than individual countries. We find that there is significant variation in the energy consumption trends within each group. The following key observations can be made for total energy consumption per capita:

- We can clearly see that low-income countries have consistently consumed below the world average levels except for Belarus.
- Most middle-income countries have enjoyed an energy usage slightly above the world average. Chile, and even more so, Argentina, have an energy usage very close to the world average.
- High-income countries have always had above-average energy consumption in the recent past. Qatar, in particular, consumes way more than any other country on the list. Singapore shows a much faster growth in energy consumption as compared to any other country on the list.

We repeat the above analysis but this time only considering renewable energy.The following key observations can be made for renewable energy consumption per capita:

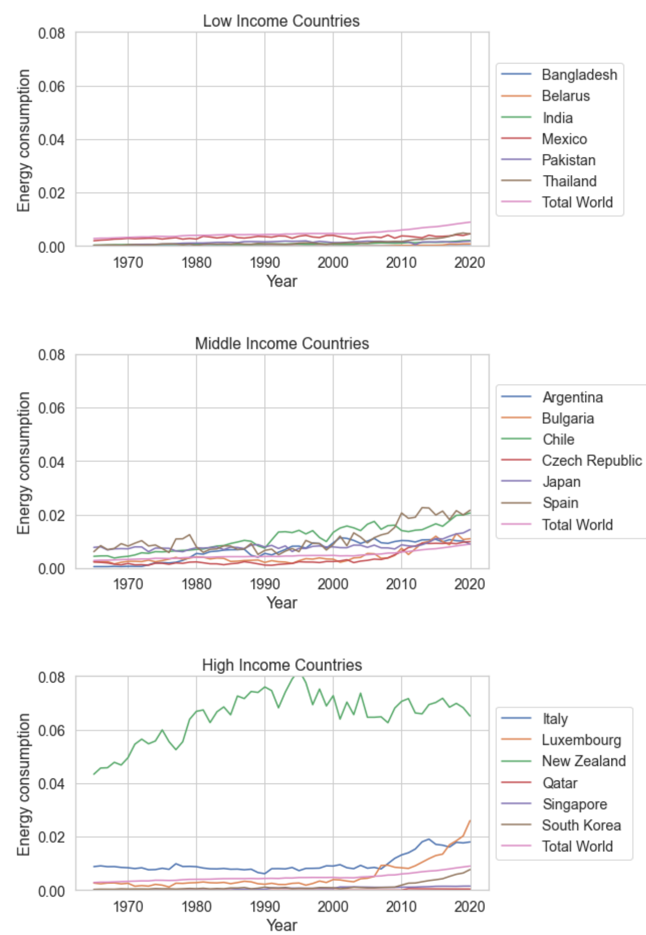

Fig. 10. Per-capita Renewable Energy Usage compared to the World Energy is in Exajoules

- We can clearly see that low-income countries have consistently consumed below the world average levels. The growth rate (as measured by slope) of renewables in these countries has been fairly low too. Thailand and Mexico have the highest renewable usage in this group

- Most middle-income countries have renewable usage well above the world average. Spain and Chile have made rapid progress in renewables and even beats many high-income countries in this regard.
- Many high-income countries including New Zealand, Italy, and Luxembourg have a renewable usage well above average. Qatar and Singapore are both notable offenders in this category. While the previous section demonstrated the rapid growth of energy usage in Singapore and high usage in Qatar, the energy usage in both cases has not been very sustainable. Luxembourg also has the highest growth rate in renewable consumption while its total energy consumption was falling. This suggests that Luxembourg has gotten better at handling energy resources efficiently.

Let us get deeper insight into each of these categories by looking at the underlying components of each energy category, that is, splitting up renewable and non-renewable sources into particular energy sources. It is clear from this analysis that higher-income countries are more likely to use renewable sources of energy. Lower-income countries largely depend on non-renewable sources. This can be explained by the high setup cost for renewables like solar and wind energy as well as non-renewable but non-fossil fuel sources like nuclear energy. High-income countries also show a decrease in coal and natural gas consumption, but a higher oil consumption.

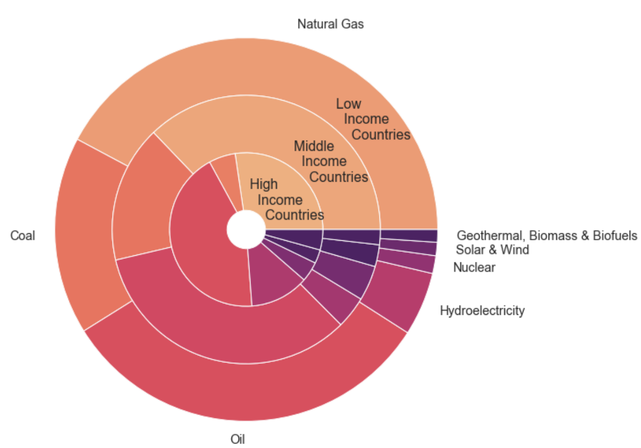

Fig. 11. Distribution of Energy Sources in Different Countries as per GDP

C. Time-evolution of energy usage for selected countries

In the above analyses we have compared the time evolution of energy usage for several countries. Now we will dig deeper and analyse a few interesting countries in more detail. We will look at the changes in relative proportions of consumption of energy sources and try to look for underlying policy or economic factors that are driving these changes in energy usage.

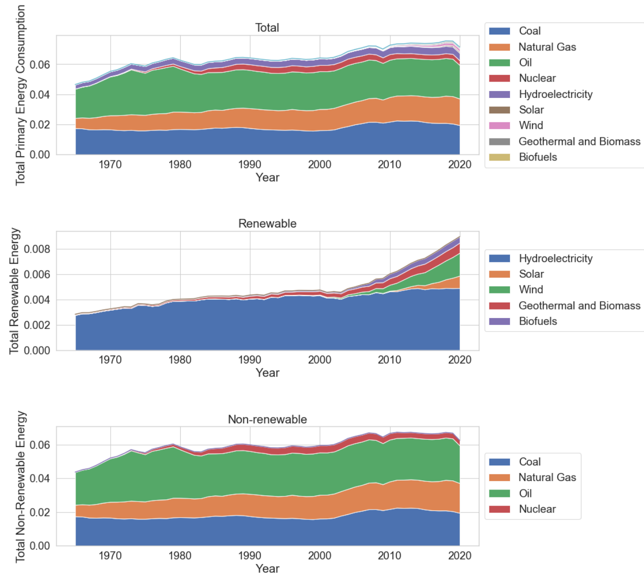

Fig. 12. Change in the Relative Share of Energy Sources (per capita) - World

World : Across the world renewables continue to play a relatively minor role. Most energy is still derived from non-renewable sources. Post 2010 we can see a rapid increase in renewable usage, particularly solar and wind energy. Hydroelectricity has always been the largest source of renewable energy across the world and its share is greater than that of all other renewables combined. The largest energy sources are oil followed by coal and then natural gas. Nuclear energy still has a smaller share than hydroelectricty but greater than all other renewables.

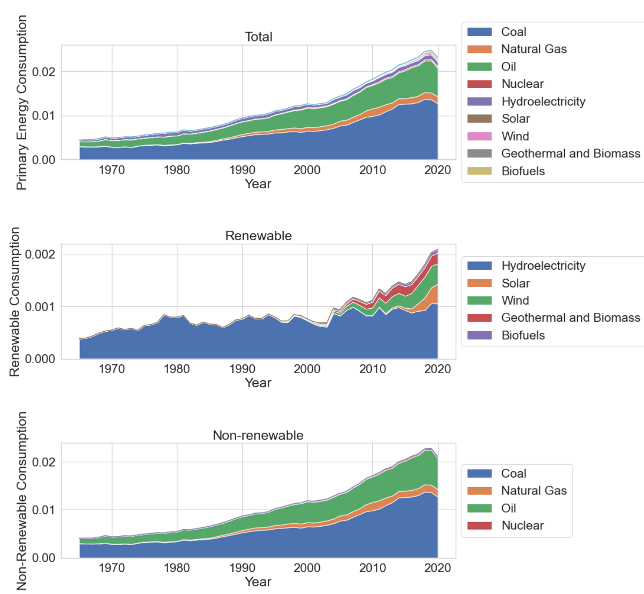

Fig. 13. Change in the Relative Share of Energy Sources (per capita) - India

India : In India, coal continues to be the largest source of energy followed by oil. Amongst renewables, hydroelectricity forms the largest portion and has a steady per capita contribution since several years. From 2004, however, we can see a rapid growth in other renewables, particularly wind and geothermal energy. Solar energy has picked up later in the country since 2012.

China : China has one of the most rapid energy growth rates. It gets most of its energy from coal which delivers more energy than all other sources combined. Post 2000 one can see a significant increase in the growth rate of per capita energy usage. China's high energy demand is not only due to the energy needs of its citizens but also the large amount of export production that it does. Most renewable energy comes from hydroelectricity.

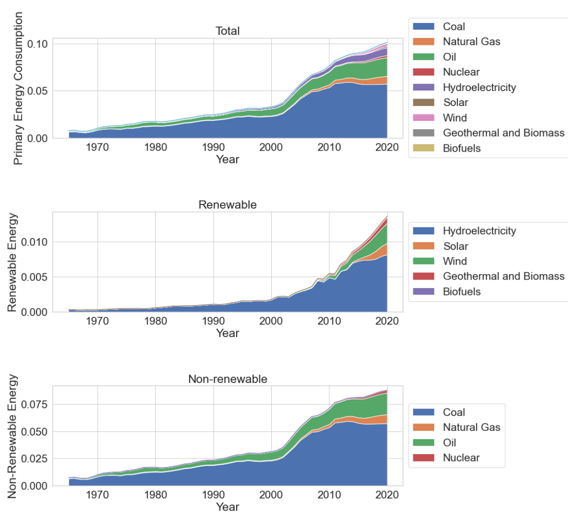

Fig. 14. Change in the Relative Share of Energy Sources (per capita) - China

Iceland : Iceland is the largest consumer of renewable energy both on a per-capita basis and by fraction of total energy consumed sourced from renewables. Unlike most countries in the world, Iceland gets most of its energy (nearly 85%) from renewables. Most of this energy comes from hydroelectricty followed by a significant share of geothermal and biomass energy. We can see a rapid increase in renewable consumption near 2008 followed by a stable level of per-capita consumption from 2010. In an era of rapid climate change necessitating a swift transition towards renewables, Iceland stands as a champion.

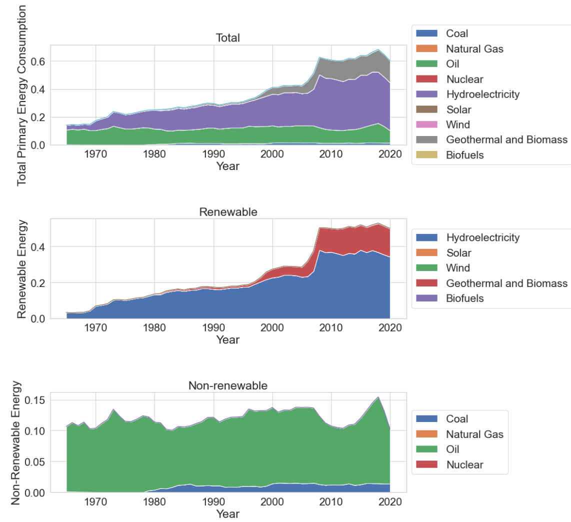

Fig. 15. Change in the Relative Share of Energy Sources (per capita) - Iceland Energy is in Exajoules

D. Effect of Policy and Events on carbon footprint

Energy usage and CO_2 emissions are complex variables that are dependent on a large number of factors. Global climate policy making platforms such as the United Nations Climate Change Conferences are major driving factors for countries to change their energy usage and greenhouse gas emissions. Besides these policy interventions, acts of god such as COVID-19 also have a major impact on energy usage.

We will focus on 4 major events in the last 20 years. These are:

- Kyoto Protocol - CMP 10 UNFCCC - International Treaty that commits to reduce CO_2 emissions - implemented from 2005
- Global Financial Crisis (2007 - 2009) - burst of US housing bubble leading to worldwide market collapse
- Paris Agreement - CMP 11 UNFCCC - International Treaty committed to keep global warming in check - implemented from 2015
- COVID-19 - Global coronavirus outbreak disrupting major supply chains and industries.

the following plots look at total quantities (rather than per-capita quantities) since we are interested in seeing effects at the global level. While no major drop in CO_2 levels is associated with either policy interventions (Kyoto Protocol and Paris Agreement), the plots show a visible dip in CO_2 levels with the global recession from 2007 to 2009 and the COVID-19 pandemic in 2019 leading until now. The dip caused by the pandemic is visible across all sectors and across most countries.

There is a visible growth in wind energy following the Kyoto Protocol and subsequently a boost in solar energy around 2015, near the adoption of the Paris Treaty. While these major global trends in energy are occurring at the same times as these global policy interventions, it is hard to say how much of this change is actually due to policy.

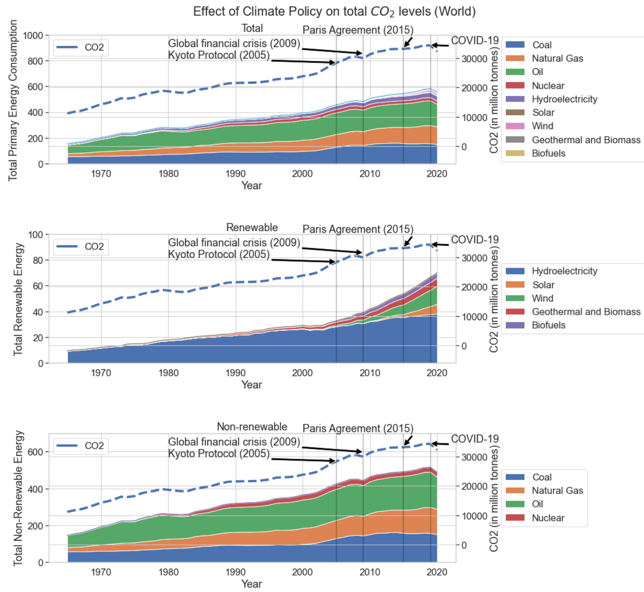

Fig. 16. Effect of Climate Policy on total CO_2 levels (World)
Energy is in Exajoules

E. *Predicting CO_2 and Energy Distribution trends*

While our analysis so far has looked at past data, we are naturally curious to understand how the energy scenario will change in the coming years. The future growth in renewables will play a crucial role in helping humanity build a sustainable future while declining CO_2 levels will help the world realise the objectives outlined in the Paris Treaty and mitigate global warming.

There are several elaborate models that draw projections for greenhouse gas emissions in the future. One of the most well known and fairly complex models is the Special Report on Emission Scenarios (SRES) by IPCC that takes into account various scenarios in global socio-economic and climate policy as well as interactions between countries.

It is difficult for us to build a model accounting for such complex global scenarios with the available data. Moreover, future trends can be strongly affected by major global events like we have seen above. Thus, it may seem that any attempt at predicting future levels is futile. But it turns out that despite such complex factors affecting these statistics, past data reveals a fairly steady change in energy usage unless a major technology development or global event takes place. To a first order, we can assume that a simple polynomial curve fitting the data will give us an idea of which sectors are expected to grow or shrink in the next few years unless a major change happens.

We aim to predict the CO_2 emissions and the distribution of energy sources in the future, and for this doing so we have proposed two models.

- Polynomial Regression followed by Support Vector Machine Regression:
  1) Polynomial Regression: For a given country and a sector, we have used polynomial regression to predict the value of expected consumption from each sector for three years. After using different degrees of the regression, degree = 3, seems the best fit.
  2) Support Vector Machine Regression: For a given year and country, we have used the sectoral data to predict the CO_2 emissions in total. As we can say that emissions majorly depend on energy consumption, we can expect a better prediction from this model.
  3) To predict the CO_2 Emissions for the next three years, we will use the polynomial regression predictions as input to our Support Vector Machine.

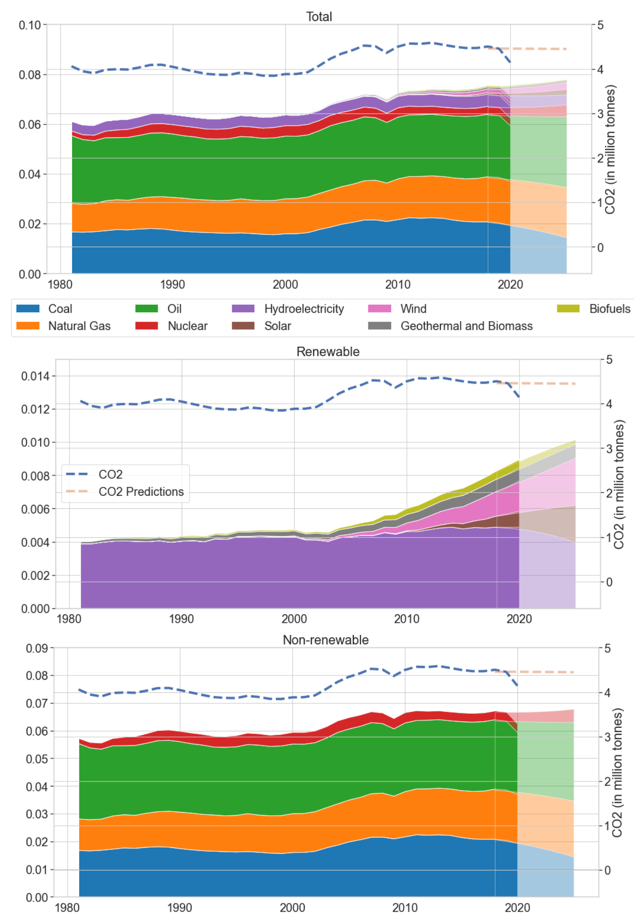

Fig. 17. Predicted Change in relative share of Energy Sources And CO_2 emissions (per capita) - World
**the predictions are plotted with a decreased opacity.*

- Polynomial Regression: In this approach, we have used polynomial regression to predict CO_2 emissions. We again observe that using a three-degree regression, the fit seems to be the best by eyeballing the plots.

We have used polynomial regression on the data of each sector for the years ranging from 2000 to 2019 and extrapolated the data for 2020 to 2025. The reason for excluding the data for 2020 is that there is a sharp and abrupt decline due to the

COVID-19 pandemic and its effect will soon be immaterial. This can be clearly seen in the plot.

The increase in the contribution of renewable energy is visible, but still, a major change is not expected unless there is some radical change in country policies which cannot be predicted by our models.

VI. Learning, Conclusions, and Future Work

A. Learning

This project exposed us to several aspects of global and domestic climate policy. We learnt about state of the art carbon emission prediction models and recognised the complexity of problems involving the climate. We also learnt that the world offers tremendous diversity in energy usage as different countries face unique constraints and have different policies and goals.

We significantly improved our grasp over the data analysis and visualisation libraries used with python, namely, matplotlib, seaborn and pandas. We also learnt the use of geopandas to visualise data on the world map and work with shapefiles. We learnt the shortcomings of machine learning techniques in predicting complex phenomenon pertaining to humans and climate and recognised that certain events are inherently unpredictable even with a lot of information.

We learnt how to collaborate over code in real-time with VSCode LiveShare. We also got a strong grasp of git and Github to manage code effectively. Working in a team taught us how to make the most of everyone's skills.

B. Conclusions

The original data had a lot of variables that were not directly useful for our analysis. We restricted ourselves to the energy consumption from those key energy sources which had sufficient data available. We also transformed the GDP and HDI data significantly to integrate it with the energy data. This data preprocessing allowed us to efficiently draw conclusions and visualise data.

There is a strong correlation between energy measurements and several other metrics like population, GDP per capita and HDI of a country. These correlations offer 3 key insights: In rich countries the average person consumes more energy, in rich countries a larger share of energy is sourced from renewables and the well-being of a country is strongly correlated with energy availability.

Despite a rapid growth in renewables in the last decade, the majority of the world still runs on non-renewable sources of energy that have a larger carbon footprint and are unsustainable. Policy-makers are faced with two conflicting goals - reducing CO_2 emissions and increasing well being - and the challenge is that these quantities are positively correlated due to our high dependence on non-renewables. This means that rapid adoption of renewables will pave the way for a better future for everyone.The exponential population growth makes this even more challenging as it is not sufficient to only increase the total energy available but also the energy available per person which increases quite slowly.

The impact of climate policy interventions in the past has not been very significant. While future projections are optimistic, because of an increasing share for renewables, they do not indicate a sustainable trend due to per capita CO_2 levels not declining fast enough. Global warming is one of the toughest challenges we face and stronger policy interventions and technological breakthroughs will be needed to achieve sustainaible growth.

C. Future Work

Our modelling for future energy developments does not incorporate the effect of policies. We can improve predictions with increased model complexity accounting for these scenarios. We can also analyse more parameters in our analysis such as other greenhouse gases and climate parameters like temperature and biodiversity.Similar to our analysis of policy interventions we can look at major technological breakthroughs in energy and their tangible effects on world energy usage and carbon footprint.

References

[1] British Petroleum Statistical Review - Energy Data Source, Consolidated dataset in panel format.

[2] UN Development Programme - Human Development Reports - GDP Per - Capita Dataset, HDI Dataset

[3] Statistical Review of World Energy 2021

[4] Intergovernmental Panel on Climate Change - Special Report on Emission scenarios by World Meteorological Organization (WMO) & United Nations Environment Programme (UNEP)