Strategic Analysis of Renewable Energy Integration and Greenhouse Gas Air Emissions

1st Komal Prakashchandra Patra
dept. of Computing
National College of Ireland
Dublin, Ireland
x22210369@student.ncirl.ie

2nd Vipin Sharma
dept. of Computing
National College of Ireland
Dublin, Ireland
x22207406@student.ncirl.ie

3rd Gaurav Gupta
dept. of Computing
National College of Ireland
Dublin, Ireland
x22212311@student.ncirl.ie

***Abstract*—This report shows the importance of electricity for the overall growth of the economy but at the same time considers the adverse effect on the environment due to the burning of fossil fuels during the production of electricity generation. It also states how important is the usage of renewable energy during electricity production and other major sectors contributing to greenhouse emissions.**

***Index Terms*—Electricity, Energy, Electricity installed capacity, Electricity Generation, Greenhouse Emission, Energy Transition, Renewable energy, Non Renewable Energy**

I. Introduction

The 2015 Paris Agreement is a worldwide commitment to take action on climate change by keeping global warming below the lower limit two degrees Celsius and greater than the pre-industrial levels. The agreement underlines the need for nations to track and reassess their climate-related goals on a regular basis. As we near 2030, this periodic assessment will be critical for assessing progress and making required modifications to fit with the larger goal of attaining carbon neutrality by 2050 [1].

A cornerstone of this effort lies in the transition of fossil fuel-based energy infrastructure, where a significant portion of electricity is currently derived from coal and natural gas. To meet the ambitious targets set for 2050, countries must embrace advanced energy infrastructure that integrates renewable energy sources such as wind turbines, solar farms, and bioenergy. Electricity, being a fundamental aspect of modern human life, plays a central role in this energy transition. It is not only vital for meeting basic needs but also for sustaining economic activities, technological advancements, and overall societal well-being. Therefore, the transformation of the energy sector becomes synonymous with ensuring a stable, reliable, and sustainable electricity supply.

Data analysis emerges as a critical tool in navigating the complexities of energy transition. Various agencies, including the International Monetary Fund (IMF), National Renewable Energy Laboratory (NREL), Energy Information Administration (EIA), Sustainable Energy Authority of Ireland (SEAI), and the International Energy Agency (IEA), play instrumental roles in collecting and disseminating valuable data. This data provides a rich and comprehensive understanding of the current state of energy systems, trends, and potential challenges.

The raw data collected by these agencies serves as the foundation for in-depth analyses that can uncover insights into the progress of energy transition goals. By employing sophisticated analytical tools and methodologies, researchers and policymakers can identify patterns, trends, and areas of improvement. Data-driven analysis is particularly effective in assessing the effectiveness of renewable energy integration, evaluating carbon reduction strategies, and understanding the dynamics of energy consumption and production.

Insights derived from data analysis contribute significantly to the identification of gaps in the current energy transition efforts. These gaps may manifest as barriers to renewable energy adoption, inefficiencies in existing systems, or challenges in meeting specific targets. Understanding these gaps is essential for refining policies, directing investments, and formulating strategies that address the root causes of obstacles to progress.

Furthermore, data analysis aids in predicting future trends and potential roadblocks, allowing for proactive decision-making. By leveraging historical data and employing predictive modeling, stakeholders can anticipate challenges and implement preemptive measures to ensure a smoother transition to a more sustainable energy landscape.

In essence, the success of the energy transition relies on the synergy between ambitious policy goals, technological innovation, and informed decision-making. Data collected by reputable agencies acts as a compass, guiding countries and organizations toward a more sustainable and carbon-neutral future. As we approach the milestone of 2030, the insights gained from data analysis will play a pivotal role in steering the course toward meeting and possibly exceeding the goals outlined in the Paris Agreement, ultimately paving the way for a more sustainable and resilient global energy system by 2050.

II. Related Work

F.R. Pazaheri et al [2] have done research in the past on renewable electricity, The research shows the analysis of increasing the use of Renewable energy sources based on electrical power production technologies. The author also shows different graphs that show the increasing use of renewable energy on the latest data of Solar energy, wind energy, geothermal and Biomass energies. It is concluded that in order to meet the high demand for electricity in the future without negatively affecting the environment, there is a need for the growth of RES-based electricity generation on a global scale. Another researcher, U. Sahin [3], looked into the total installed capacity of renewable and hydro energy in Turkey as well as the country's electricity generation from 2019 to 2030. In 2030, it was projected that there will be 80.3 GW of installed renewable capacity and 241 TWh of electricity generated. Turkey's installed hydropower capacity and electricity generation are expected to reach 30.7 GW and 57.3 TWh, respectively, by 2030, according to the Nonlinear Grey Bernoulli model. The FANGBM model predicts that in 2030, hydropower will account for 38.2% of all installed renewable capacity and 23.8% of all electricity generated. Global trends in greenhouse gas emissions are displayed in another study by J.G.J. Oliver and J.A.H.W. Peters [4]. Emissions in 2019 totaled 57.4 GTCo2, which is an approximate 70% increase from 2018. This increase took place in 2019 as the world economy grew by 2.8%, a rate that is marginally slower than the 3.7% and 3.5% global GDP growth in 2017 and 2018, respectively. Sam Wood and Annette Cowie [5] discussed that the production of fertilizers (nitrogen concentrated) contributes high to greenhouse gas emissions during electricity generation, primarily through CO2 from natural gas combustion during ammonia synthesis and N2O during nitric acid production. D weisser [6] discusses about the ways to reduce greenhouse gas emissions from electricity generation, accounts for upstream and downstream emissions, promotes efficient fossil fuel usage, transitions to low-carbon fuels, expands to renewable and nuclear energy, and develops carbon capture technologies to reduce global emission. William R. Patterson [7] talks about the key gases—carbon dioxide (CO2), methane (CH4), nitrous oxide (N2O), and fluorinated gases. As well as discusses about the sea-level rise, extreme weather events, and resource conflicts.

III. Methodology

A. Technologies Used:

- Python: Using the Python package we show the number of different interactive charts which is helpful for the analyzing data.
- Visual Code Studio (VS Code) : It is a popular Integrated Development Integrated (IDE) developed by Microsoft and we use this for performing Exploratory Data Analysis (EDA)
- PostgreSQL: The Postgres database is the open-source relational database management system commonly used for scalability, and multi-version control to have all checkpoints. Committed.
- MongoDB: MongoDB is an open-source NoSQL database where the data is in document-based structures. It was used to store un-structured data like XML and JSON.
- Dagster: It is used to create a Pipeline of the entire process and it is also an open-source data tool designed to help build, deploy, and monitor data workflows.

B. Python Packages Used:

- Data Pre-Processing: requests, csv, Numpy, Pandas, json, dagster
- Data Storage and Retrieval: Psycopg2, pymongo, sqlalchemy
- Data Visualization: Matplotlib, Seaborn, plotly

C. Data Gathering

1) Dataset 1 - Electricity Generation data from 2000 to 2021:

Electricity Generation defines the total electricity generated by different countries using different technologies like Bioenergy, Fossil fuels, geothermal, Hydropower, Solar energy, wind energy, Nuclear, etc. This electricity generation data shows the yearly electricity generated from 2000 to 2021. The data mainly focuses on type of Energy which are Renewable and Non-renewable. The Renewable Energy that comes from clean energy which naturally occurs from wind, solar, marine or Bioenergy, etc., on the other hand Energy that comes from Nuclear, Pumped Storage, Fossil fuels including coal, crude oil, natural gas, and uranium is commonly referred to as Non-Renewable energy. The overall electricity produced by the CHP (Combined heat and power) plants, Power plants and the distribution generators are measured as the output terminals. It comprises of both off-grid and on-grid electricity generation as well as the power utilized by the other technology. The unit of Electricity Generation for this dataset is Gigawatt-hours (GWh). The data has been collected as a semi-structured JSON format using the API from IMF Climate change dashboard. The dataset consists of 1000 rows with 25 attributes.

2) Dataset 2 - Electricity Installed Capacity data from 2000 to 2021:

The Installed Capacity of an Electricity defines the highest possible Electricity generation that the installation can provide and is generally expressed in Megawatts (MW). The maximum active power which is supplied continuously during the whole period where the plant keeps on running. Electricity installed capacity data define the installed capacity across the various technologies like Bioenergy, Fossil fuels, geothermal, Hydropower, Solar energy, wind energy, and Nuclear in the different countries same as the electricity generation data. For analyzing the different countries installed capacity to meet their electricity needs, installed capacity is a very important metric and it is important to know that the installed capacity does not give the amount of electricity produced at any given moment in time. Many agencies and the government use the data of installed capacity to make decisions about the consumption of electricity in different regions or countries. This data was also gathered by the Climate change dashboard.

3) Dataset 3 - Annual Greenhouse Gas Emission from 2010 to 2021:

The annual greenhouse gas emissions dataset consists of emissions from the sum of carbon dioxide (CO2), Fluorinated gases, Methane and Nitrous oxide. They are measured in Million metric tons of CO2 equivalent. The data also segregates the emission based on different Industry such as Agriculture, Forestry and Fishing, Construction, Electricity, Gas, Steam and Air Conditioning Supply, Manufacturing, and mining. Apart from this, division is based on continent-wise, sub-continent wise, Advanced Economies and Emerging and Developing Economies for better classification. The dataset is fetched from API and saved as the structured data as CSV. The dataset consists of 1140 rows and 24 columns.

D. Data Management:

For managing and analysing the data, the researcher has used various tools such as databases like PostgreSQL and MongoDB, programming languages like Python and ETL tool as Dagster.

PostgreSQL :

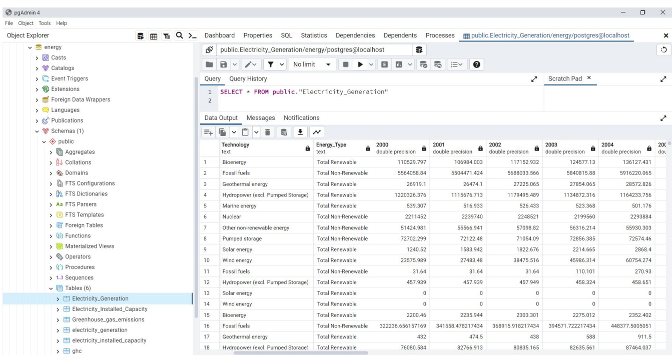

Fig. 1. PostgreSQL

The Postgres database is the open-source relational database management system commonly used for scalability, and multi-version control to have all checkpoints committed, robust, and supports numerous programming languages such as C++, Java, C, Python, etc. PostgreSQL can handle complex structured data efficiently. As it also supports ACID properties (Atomicity, Consistency, Isolation, Durability).

This database has been used to store the structured data that we have as a comma separated file fetched from API. The main objective of using this dataset is to create a centralized database system from where we can fetch the data as per the requirements and visualize to get useful insights.

MongoDB :

MongoDB is an open-source NoSQL database where the data is in document-based structures. Semi-structured data such as JSON and XML which is in document format can be stored in this database. MongoDB is a distributed database where they are built on top of cloud environments like Azure, AWS, and GCP. Each document has a flexible schema which means every document can have a different field as there is no restriction of having the same schema, as they are schema-less. Every document has a unique key within a collection and can also be nested, as they are hierarchal. It supports BASE properties and CRUD operations (Create, Read, Update, and Delete). This database has been used to store semi-structured data as a document because the volume of data varies from each other, and they are stored in the RAM for fast and better performance. Two of the datasets fetched from API as a JSON format have been stored in MongoDB.

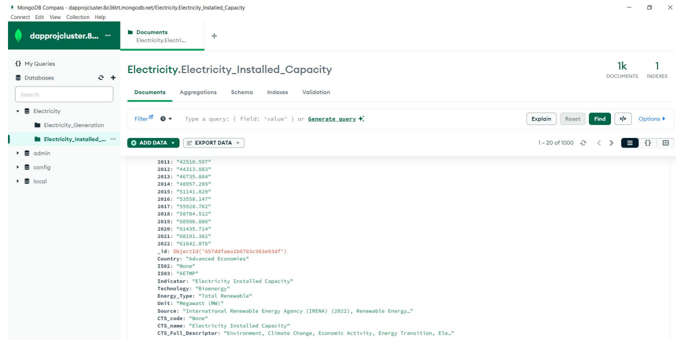

Fig. 2. MongoDB

Dagster :

Dagster is a data pipeline orchestrator tool that automatically performs a sequence of processes and keeps track of results for data integration, data validation and monitoring. Using Dagster, assets in it can model data produced by any system, such as dbt models, snowflake tables, and so on. The assets can be developed both with and without the use of software defined assets. With software-defined Assets, the version control can be built (mostly the data/datasets are versioned) and is primarily used for the complex data pipelines while Ops and Jobs are used for without-software defined assets and are used for simple data pipeline. They also use pipelines for encapsulation, code reusability, robustness, and maintainable data workflows.

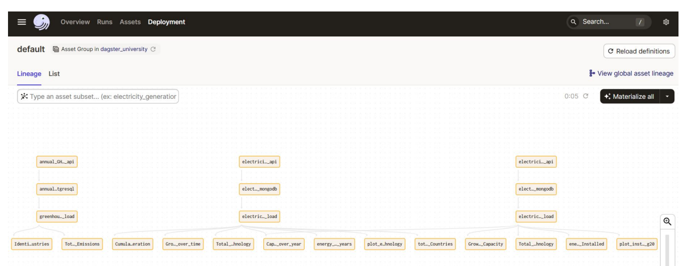

Fig. 3. Dagster

GitHub :

GitHub is a collaborative software development tool used for continuous integration and continuous development. The main aim is to manage the code repository and track changes in code. It also allows users to create public and private repositories. As it supports version control, integration with various tools makes it famous among developers. The researchers have used GitHub to manage their project version. The initial step is to create a new repository on the GitHub platform for creating a new project which involves the project name, description, and repository visibility (Public or private) and optionally initialize the README file. Once the repository is set up properly then It can be used by several users/ members of team to clone to their local machines (Visual Studio IDE) using the repository URL and can keep on pulling and pushing the codes by committing them with a small message to maintain the proper log about the version.

E. System Flow and Data Processing Implementation

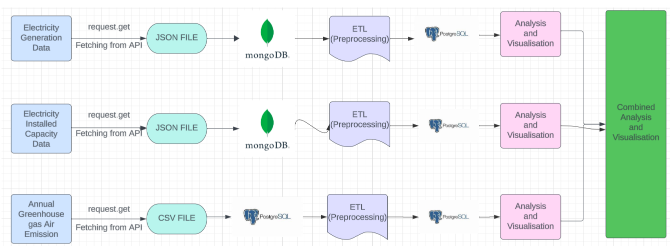

Fig. 4. SystemFlow Diagram

The Three datasets that have been used for this research study are Electricity Generation, Electricity Installed Capacity and Annual Greenhouse gas Air emission from IMF Climate change Dashboard. Using request, the data was retrieved from API (Application Programming Interface) where two datasets are acquired and stored as JSON files, which is a key-value pair and the third one as structured CSV file.

The JSON files are stored in the NoSQL database 'MongoDB', this database has been selected for its scalability and schema-less. Because the documents are independent of each other, they are not affected by the size of the document, which results in better and faster query performance.

Once all the raw data are stored in their respective databases, the Extract, Transform and Load or ETL process has been taken place in which JSON data is retrieved from MongoDB and CSV data from PostgreSQL to analyze and propose solutions for the research objectives. In order to obtain the accurate and clean data as well as to boost overall productivity, The datasets has been cleansed to get the high quality and relevant data. There are several steps we perform in the cleaning like handling the missing values, in our datasets there is no missing values, checking the datatypes of the variables, etc. The unwanted columns such as IS02, IS03, Indicator, Unit, Source, CTS code, CTS name and CTS full descriptor are dropped from all three datasets. The column names are transformed to new names so that the columns can be easily understood and interpreted. The datasets do not have any missing values, but there are values which are null or None, which is then filled with 0 as well as the numerical data which had String datatpe are converted into numerical column. Once the transformation for all the three datasets has been performed, they are loaded back to databases. This time all the datasets are loaded back to the PostgreSQL database in a structural format. All this has been done by using Python. Several libraries of Pythons such as pandas, pymongo, psycopg2, requests, matplotlib, seaborn, sqlachemy, io.sql, etc. are utilized to fetch, perform ETL operations as well as Visualisation.

All this process is orchestrated as a data pipeline to automate within a single process of control flow using Dagster. The Software-defined assets are used in this research study in order to manage the complex data pipelines as it is bit complex. These assets are the fundamental block of dagster which is materialized by which the pipeline that asset has been updated or created can relate to a dependency. They are the relationship between the two assets which can be represented by upstream and Downstream. The assets are created for each function in order to perform in a single flow. All the configurations and credentials are stored in a Resource. All this has been scheduled for executing the task to kick off runs without human intervention.

IV. Results and Evaluation

The results are evaluated for this research study fetched from the PostgreSQL database and stored in pandas DataFrame to further analyze the data by using various types of visualisations. All the visualizations are performed independently in order to get the insights that can be interpreted. They are performed using Python libraries such as matplotlib and seaborn.

A. Total Electricity Generation over the years for Different technologies:

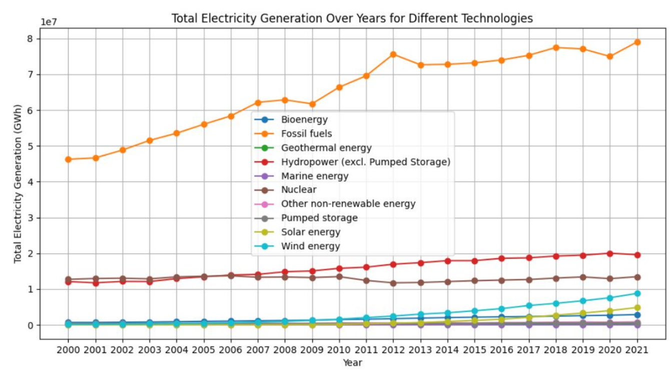

Fig. 5. Electricity Generation

The above fig 5,6 represents the Total Electricity Generation and Total Electricity Installed Capacity over the two decades for various technologies. Fossil fuel has been the dominant source for generation as well as installed capacity. However, there are trends towards cleaner energy due to environmental impact and the aim of all countries to reduce greenhouse carbon emissions by 2030. Renewable energy such as Hydropower, wind, solar, and Pumped storage has started growing consistently due to the installation of PV systems, and an increase in wind farms even though the rise is slow as compared to non-renewable resources still it contributes to clean energy. Nuclear power has been proven as a reliable and stable source of electricity but safety concerns and the increase in cost have hampered growth. The growth of wind and solar

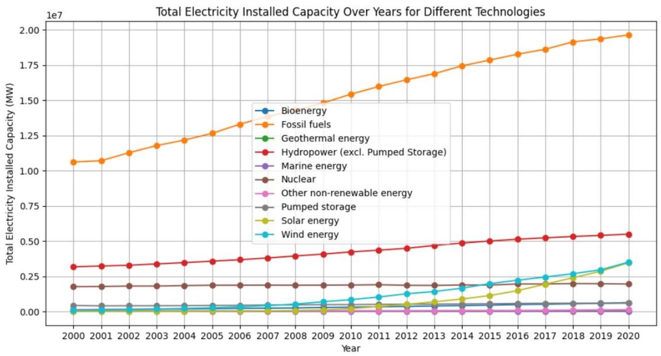

Fig. 6. Electricity Installed Capacity

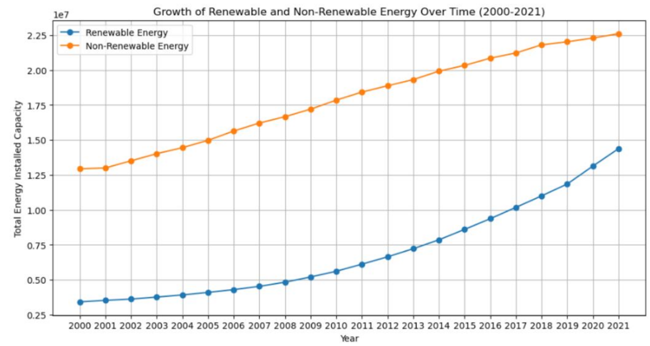

Fig. 8. Growth of installed Renewable and Non-Renewable

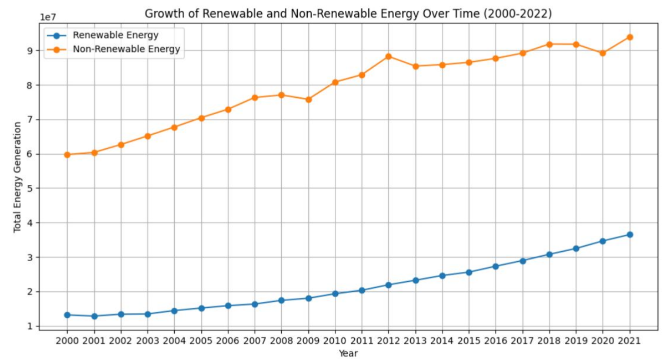

Fig. 9. Growth of installed Renewable and Non-Renewable

energy in the case of Installed capacity has been significantly growing from 2010.

B. Cumulative Percentage share of technologies for Total Electricity Generation:

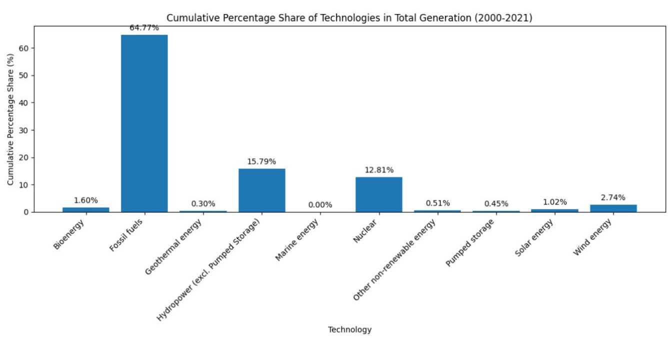

Fig. 7. Contribution of Technology in Total Electricity Generation

Percentage share is the total electricity generated for each technology by total Electricity generation. The above fig 7 illustrates the 64.77% contribution of Fossil fuel being the dominant source to generate electricity globally. As Nuclear power offers a low carbon energy that does not produce greenhouse gas emissions while Hydropower is already being a renewable and clean energy source, a very high cumulative proportion is contributing towards lowering the carbon footprint. Pumped storage and Geothermal are commonly regarded as dispatchable energy sources, in which Pumped storage helps in preserving and discharging energy on demand,while geothermal energy is a constant and reliable source. They have considerable upfront expenses that may have affected the contribution.

C. Growth of Renewable and Non-Renewable over years:

The below fig 8,9 is the analysis of growth in types of energy such as renewable and non-renewable over the years. The renewable energy is significantly experiencing growth globally where governments of all countries are targeting to invest in several renewable projects to diversify the energy mix and reduce carbon emissions. Renewable energy has been rising from 18.1% to 28% contributing of total 21.5% in the case of electricity Generation in the last two decades while Non-renewable energy have decreased from 82.5% to 72% contributing to an overall fall of 10.5%. Therefore, the Global renewable electricity capacity is predicted twice from 2020 levels by 2026 and will be tending towards 4,800 Gigawatt.

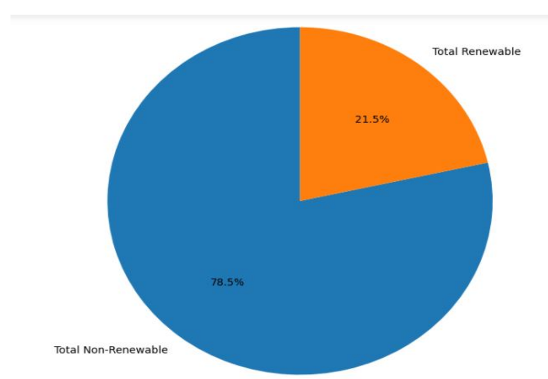

Fig. 10. Contribution of Renewable and Non-Renewable Energy

D. Renewable and Non-renewable share in the total energy mix:

The Fig 11, 12 illustrates that the share of Renewable energy for both Electricity Generation as well as Installed Capacity is increasing slightly due to which the effect on non-renewable energy is decreasing because of initiatives to minimize carbon emissions and switch to better energy sources. Energy mix refers to the collection of different energy sources utilized to suit the demands of a certain geographic location.

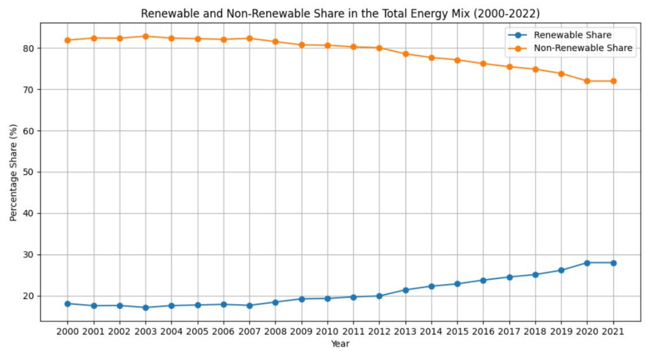

Fig. 11. Renewable and non renewable share in total energy

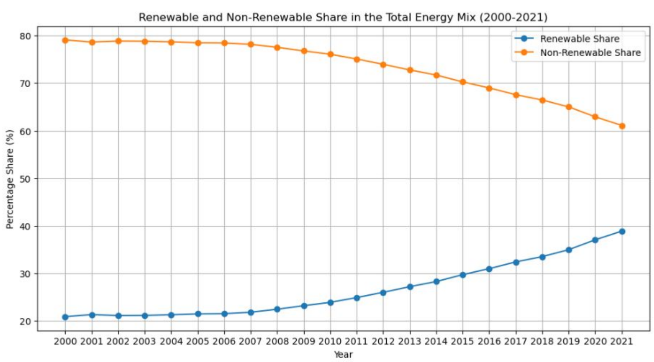

Fig. 12. Renewable and Non Renewable share in total 2000-2001

E. Capacity Factors by Technologies:

The ratio of the electrical energy generated by a producing unit over a certain time period to the electrical energy that might have been produced during the same time period if the unit had been running at full power continuously. The fig 13 illustrates to analysis of the capacity factor with respect to technology where the capacity factor defines conversion technology and its significance. The capacity factor for Geothermal power generation is the highest among all other energy conversion processes due to its non-fuel cycle, continuous heat availability, and constant power output assuring stable electricity generation to the grid. Knowing a technology's mean CF gives crucial details for designing and expanding the infrastructure necessary for supplying power.

F. Total Greenhouse Gas Emissions Over Years by Gas Type:

As data involves around time series, so constructed a line chart of total greenhouse gas emissions over the year in million metric tons by gas type. It shows that Fluorinated gas contributes the least to the emissions. Carbon dioxide and Greenhouse gases are the main sources of greenhouse gases.

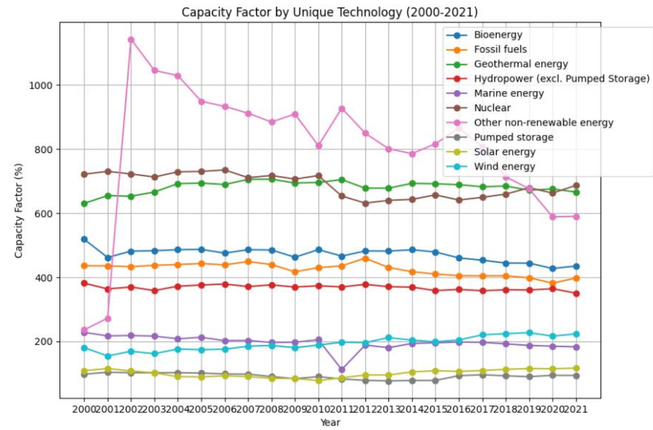

Fig. 13. Capacity Factors by Technologies

The contribution of carbon dioxide is around four times that of Methane. Also, it shows that both these gas emissions have been increasing at a constant rate for the last 10 years Fig 14

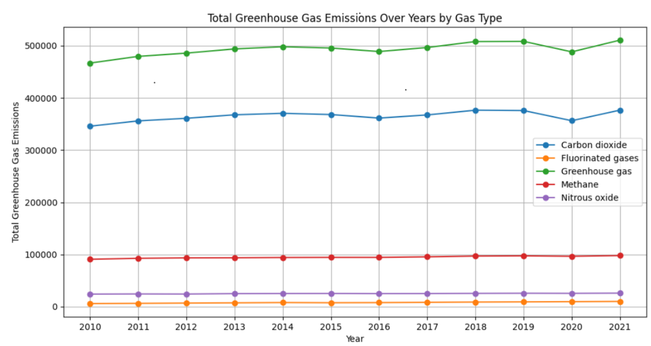

Fig. 14. Greenhouse Gas Emissions Over Years

Funnel chart in Fig 15 graph shows that Greenhouse gas gases contribute 49% of the total of emissions. The contribution of carbon dioxide and methane is 4.38 million and 1.13 metric tons, respectively. This is because C02 has higher persistence in the atmosphere than any other gas. Secondly, C02 is the major generated during the combustion of fossil fuels such as electricity generation and industrial activities.

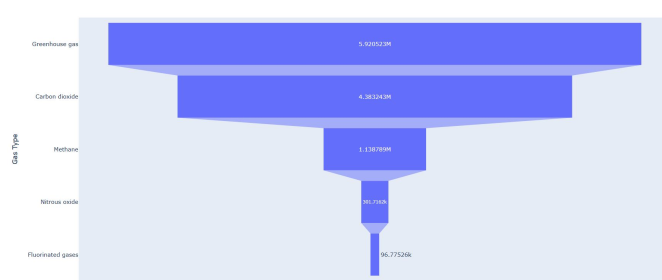

Fig. 15. Contribution of each Gas type

G. Total Greenhouse Gas Emissions by Industry:

The bar chart in Fig 16 shows the emission of greenhouse gases sector-wise, which shows total industry and household, and electricity generation are the two major causes of greenhouse gas emission. This contribution is greater than the summation of all other sectors. In fact, Mining, Agriculture, and Manufacturing also contribute less than Electricity Generation. This is due to the high dependence on fossil fuels during the Electricity generation, and there are limited clean energy and efficient technologies.

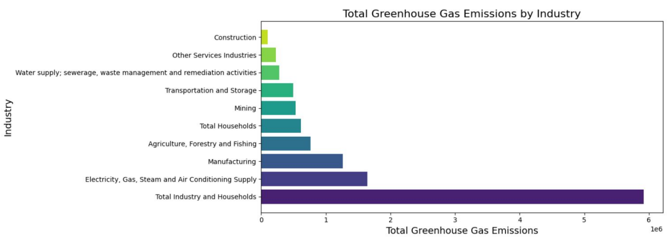

Fig. 16. Greenhouse Gas Emissions Sector-wise

H. Mean Greenhouse gas emission by countries:

The line chart in Fig 17 of mean emissions by countries shows that Asian countries are the major source of emissions (mostly South Asian countries). In fact, Advance Economies and Emerging and Developing Countries' contribution is more than the rest of the countries. It can be interpreted that this is due to improper rules and regulations. Their primary aim is to earn profit rather than using expensive cleaner technologies.

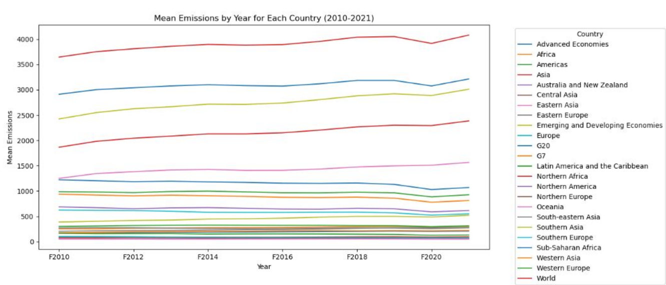

Fig. 17. Mean Greenhouse gas emission Country-wise

V. Conclusions and Future Work

In this research work, we conducted a comprehensive analysis of electricity generation, electricity installed capacity, and Annual greenhouse gas emissions. Electricity generation data provided valuable insights that involve the landscape of global energy production and its environmental implications. The fossil fuels in both electricity generation and installed capacity show the need for a planned change to more sustainable renewable energy sources. Based on the visualization of greenhouse gases, we analyzed the insights of each gas and their contribution to greenhouse emissions and considered the contribution of sector-wise. As well as getting insights based on country-wise, developing, and emerging economies.

In the future, we can use time series algorithms to make a prediction of electricity generation and greenhouse gas emissions so we can formulate policies and technologies according to that respect. As seen in this paper, barriers in advanced technologies limit the efficiency of electricity production and increase the emission of such gases, so introducing modern technology and their effect on the overall scenario. In-depth analysis of each factor to optimize electricity generation and greenhouse emissions at the same time.

References

[1] https://www.iea.org/news/renewable-electricity-growth-is-accelerating-faster-than-ever-worldwide-supporting-the-emergence-of-the-new-global-energy-economy.

[2] F. R. Pazheri, M. F. Othman, and N. H. Malik, "A review on global renewable electricity scenario," Renewable Sustainable Energy Reviews, Mar. 01, 2014. https://doi.org/10.1016/j.rser.2013.12.020.

[3] U. Şahin, "Projections of Turkey's electricity generation and installed capacity from total renewable and hydro energy using fractional nonlinear grey Bernoulli model and its reduced forms," Sustainable Production and Consumption, Jul. 01, 2020. https://doi.org/10.1016/j.spc.2020.04.004.

[4] J.G.J. Oliver and J.A.H.W Peters, "Trends In Global CO2 and Total Greenhouse Gas Emissions," December 2020. [Online]. Available: https://klimaatweb.nl/wp-content/uploads/po-assets/372210.pdf.

[5] Sam Wood and Annette Cowie, "A Review of Greenhouse Gas Emission Factors for Fertiliser Production", IEA Bioenergy, June 2004.

[6] Daniel Weisser, "A guide to life-cycle greenhouse gas (GHG) emissions from electric supply technologies", International Atomic Energy Agency, 31 May 2006.

[7] William R. Patterson, "Greenhouse Gas Emissions", Science Advisory Group of the United Nations (UN) Climate Action Summit, 2019.