Energy Pulse: Global Trends in Power and Emissions

Team 9:

- Pawan Bath - 30067612
- Vladimir Stepanov - 30160874
- Yuhan Zhang - 30227385
- Vhanessa Cardona - 30261993
- Malhar Dhawle - 30238470

1. Introduction

1.1. A clear motivation for the data exploration

Energy growth is directly correlated to prosperity and well-being globally. Energy demand has increased rapidly to support economic and social progress, particularly in developing countries. It is estimated that globally around 750 million people – 1 out of 10 – do not have access to electricity and around 2.6 billion people rely on heavily polluting biomass fuels such as charcoal, coal, and animal waste for heating and cooking (Energy Institute, 2024). According to the UN, the transformation and use of energy is the single biggest contributor to global warming. The energy sector, dominated by fossil fuels, accounts for 34 percent of human-caused greenhouse gas (GHG) emissions, with a total of 20 Gigatons (Gt) of GHG globally (IPCC, 2023).

To better understand these challenges, our project investigates global energy trends, focusing on how different energy sources—both fossil-based and renewable—contribute to CO_2 emissions. By analyzing patterns in energy production and consumption, our work aims to highlight critical areas for intervention and provide insights into the pathways toward a more sustainable energy future. This study builds on the broader context of balancing energy equity with environmental responsibility, connecting global trends to actionable solutions.

1.2. Research questions

- What are the general energy trends in carbon emissions?
- What is the relationship between renewable energy capacity and carbon emissions?
- How do the countries with the largest oil trade volumes contribute to global carbon emissions?
- Is there a noticeable impact on electricity generation fuel mixes in countries with significant renewable energy installations?
- How do countries heavily involved in oil imports or exports compare in their reliance on oil for electricity generation?

1.3. Background knowledge important for understanding the data exploration

Understanding the relationship between energy generation, trade, and carbon emissions is critical for analyzing global energy trends. Renewable energy sources, such as wind and solar, are essential for reducing carbon emissions, as they generate electricity without relying on fossil fuels (Tierney, n.d.). In contrast, traditional energy systems heavily depend on coal, natural gas, and oil, which are among the largest contributors to global carbon dioxide emissions. The global oil trade further highlights the dependency on fossil fuels, influencing both energy production and carbon footprints. Variations in electricity generation fuel mixes across countries, often shaped by policy, resource availability, and renewable energy capacity, provide valuable insights into energy transition efforts (*Executive Summary – World Energy Outlook 2024 – Analysis - IEA*, n.d.). By examining installed renewable capacities, fuel mixes, and trade data, this project explores trends and relationships to better understand carbon emission patterns and the role of renewables in achieving sustainability.

2. Datasets

2.1. Source and licensing of the dataset

“World Energy Consumption” data collection by Pralabh Poudel on Kaggle (link) or alternatively Github (link) served as a great start for the team to initiate the analysis and start collecting the required information, modify the data, provided proper attribution is given. A full disclaimer on liability, terms of use, reservation of rights and indemnity can be found by following the link .

Though, identified collection of information on Kaggle contained substantial amount of data for Primary Energy consumption and Renewable Energy, it did not contain information on the carbon dioxide (CO2) emissions and global trade of petroleum products, which are critical for our analysis. Referring to the resources mentioned on the Kaggle page, we’ve identified that Energy Institute held the collection of the required data in the form of excel file (link) with 81 individual tabs covering information on various energy sources by country and region.

2.2. Dataset distribution

The distribution of the datasets is as follows:

- Renewables - Generation by source (Malhar Dhawle)
- Electricity generation by fuel (Pawan Bath)
- Oil: Trade 2022 and 2023 (Vladimir Stepanov)
- Carbon: Carbon Dioxide Emissions from Energy (Yuhan Zhang)
- Renewable Energy - Solar and Wind Installed capacity by type (Vhanessa Cardona)

2.3. Structure and format of each dataset (**for each member**)

2.3.1. Renewables - Generation by source

This dataset is a collection of renewable energy data by source and country. It provides a detailed view of renewable energy production and consumption across multiple countries worldwide for 2022 and 2023. This data is formatted where each row represents a country and each column represents a renewable energy source and the year. The dataset contains the following columns:

- Wind(TWh): Generation from wind.
- Solar(TWh): Generation from solar/photovoltaic.
- Hydro(TWh): Generation from hydroelectric sources.
- Other Renewables(TWh): Other renewable sources include but are not limited to geothermal and biomass.

The generation values are in TWh (Terawatt hours) which is a standard unit used for measuring large-scale energy production. This dataset allows us to perform regional comparisons where we can compare renewable energy production across different countries. It also enables us to perform source-specific analysis to compare the production levels of different renewable energy sources. It helped us understand global trends, patterns and disparities in renewable energy production. It also gave us an idea of potential technological advancements and policy changes across different nations.

2.3.2. Electricity generation by fuel

This dataset analyzes electricity generation by fuel measured for 2022 and 2023 in 37 countries. The different sources are oil, natural gas, coal, nuclear energy and hydroelectricity. The dataset also includes columns for "Other Renewables" and a generalized "Other" category which includes uncategorized generation, statistical differences and sources not specified elsewhere e.g. pumped hydro, non-renewable waste and heat from chemical sources. Aggregate totals for each region and energy type are also included. The fuel is measured in TWh (terawatt hours).

This dataset provides insight into energy production trends, which is vital for analyzing the global transition towards renewables, regional reliance on specific energy sources, and to analyze trends in certain fuel sources.

2.3.3. Oil: Trade 2022 and 2023

This dataset represents an overview of the global oil trade in 2022 and 2023 for four different variables – crude oil imports, crude oil exports, refined product imports, and refined product exports. The data is arranged by country or region, which makes it possible to see in greater detail oil trading trends over these two years. For each country, the dataset contains the sum totals of crude and refined products exchanged, enabling comparisons from year to year and trends.

The format of the data also lends itself to analysing trade balances, defining major exporters and importers, and predicting how oil trade fluctuates across time. By displaying crude imports and exports from 2022 to 2023 relative to the previous year, the data gives an insight into changes in demand, production, and export capacities between countries. This makes it possible to identify areas where oil trade is increasing or declining, which reflects changing energy trends around the world. Moreover, the detailed breakdown of trading the crude and refined products demonstrates the role each country plays in the extraction, processing and transportation of the resources of oil.

2.3.4. Carbon: Carbon Dioxide Emissions from Energy

This dataset is structured in a tabular format, providing carbon dioxide emissions data for various countries from 1965 to 2023. The first column lists the names of the countries, while each subsequent column represents annual emissions data for each year within the specified range. Each row corresponds to a specific country, allowing for straightforward comparisons of carbon emissions over time. The emissions values are numerical, indicating carbon emissions in million tonnes. Some entries contain zero values, which could signify no recorded emissions for certain years.

The dataset encompasses a diverse range of geographical entities, including individual countries and seven distinct regions from around the world. This structure allows for comprehensive analysis and comparison of carbon emissions across various global contexts, facilitating insights into regional trends and patterns in emissions data. Overall, this dataset provides valuable insights into the evolution of carbon emissions in relation to various factors such as oil trade and energy consumption.

2.3.5. Renewable Energy - Solar and Wind Installed capacity by type

This dataset is a single sheet within a larger Excel file and comprises two primary tables: one dedicated to solar capacity and the other focused on wind capacity. Both tables share a common column, Country, which serves as a point of overlap. However, not all countries are represented in both tables. While there is some overlap, many countries appear exclusively in either the solar or the wind capacity table.

The solar capacity table provides detailed data about two types of solar energy production: photovoltaic capacity and concentrated solar power. For both categories, information is available for the years 2022 and 2023, allowing for a comparison of trends or changes over this period.

On the other hand, the wind capacity table focuses on two distinct types of wind energy production: onshore wind capacity and offshore wind capacity. Similarly to the solar table, data for wind energy is also presented for the years 2022 and 2023, offering insights into the development of onshore and offshore wind resources over these two years.

2.4. Noteworthy challenges (**for each member**)

2.4.1. Renewables - Generation by source

This dataset provides a detailed view of renewable energy production across multiple countries. Each column represents a renewable energy source (Wind, Solar, Hydro, or Other Renewable), and each row corresponds to a specific country. While the data was well-structured, a few challenges had to be addressed before performing any analysis.

The dataset included metadata such as regionally aggregated, global summary, and world total values, which were not relevant for analysis and required filtering out. Some cells contained placeholders (e.g., '^' indicating values less than 0.005) or had missing values, which needed to be handled appropriately. Additionally, ensuring consistent power units across all columns was essential for reliable analysis. Finally, columns were renamed and reorganized using the pandas library to make the dataset easier to load into an SQL database and prepare for further analysis.

2.4.2. Electricity generation by fuel

This dataset provides an overview of electricity generation by fuel source. There were some noteworthy challenges when it came to preparing the dataset. There were placeholders (e.g., '^' indicating values less than 0.005) and missing data throughout the dataset. Missing and incomplete data can skew the outcomes if not handled appropriately.

Inconsistency with column headers would cause issues in SQL analysis so they needed to be flattened and standardized. The columns with “Other” and “Other Renewable” lacked context and relevance so deciding how to handle that data was crucial to ensuring an accurate conclusion.

2.4.3. Oil: Trade 2022 and 2023

While oil trade dataset is quite comprehensive, there are still several challenges that are arising while working with it.

Potential data inconsistency between countries and regions is one of the biggest challenges. Reporting thresholds can vary dramatically between crude oil and refined product imports and exports, creating differences in measurement units, frequency of reporting, or category. As an example, some countries could count crude and refined product numbers together under the same heading, and others separately. Such inconsistency makes direct regional comparisons quite challenging and creates confusion in global trade patterns. Incomplete or inaccurate data for smaller trading areas can also blur the picture, creating discrepancies that prevent a complete picture of oil trade dynamics. Third-party intermediaries and re-exports do further complicate the accuracy of the data. It is quite common now that oil is sold through intermediary hubs what may mislead the analysis and bring uncertainty in the numbers as product’s true origin or destination might not always be captured. Another noteworthy challenge is the time span of the dataset. Oil markets are particularly sensitive to geopolitical developments, trade sanctions and market shocks in the world markets that can dramatically impact the influx and flow of goods while looked at different scales. However, a dataset restricted to two years would not be able to capture and represent these changes and thus would not be able to show a full picture of trends in trade.

2.4.4. Carbon: Carbon Dioxide Emissions from Energy

The dataset is in a pivoted format, with years as columns and countries as rows. This makes it complex to create queries for trends over time and across different countries. Additionally, non-descriptive headers and metadata rows need to be cleaned up to ensure meaningful column names and relevant data are retained.

Finally, the dataset includes both granular and aggregated information, such as global emission summaries and regional data, within the same table. This mixed content complicates filtering and requires additional steps to separate and analyze specific data subsets. Without these preparations, SQL queries might produce misleading insights or fail to meet analytical objectives. Addressing these issues is essential for effective SQL querying and analysis.

2.4.5. Renewable Energy - Solar and Wind Installed capacity by type

For this dataset, one particularly noteworthy challenge lies in effectively addressing missing data. Ensuring that gaps in the dataset are handled appropriately is critical for maintaining the integrity and reliability of the analysis. This may involve strategies such as imputation, removing incomplete records, or leveraging external data sources to fill in the blanks, each of which comes with its own set of considerations and potential trade-offs.

Another significant challenge involves transforming the data into a unified format. This process requires consolidating the two tables into a single, cohesive table suitable for subsequent upload to the database. This step may include aligning columns, standardizing data formats, reconciling discrepancies between the original tables, and possibly aggregating or reshaping the data to meet the database's requirements.

3. Data Exploration

3.1. Cleaning datasets (**for each member**)

3.1.1. Renewables - Generation by source

To clean this dataset, I began by removing irrelevant rows that contained metadata. These included rows with regionally aggregated data, global summaries, and textual information that were not relevant for analysis and could cause issues during data loading into an SQL database. Placeholder values, such as "^" (indicating less than 0.005 TWh of energy generation), were replaced with zeros to ensure consistency.

Next, I performed data normalization by flattening the multi-level column names. The original dataset had higher-level headers representing the year and lower-level headers indicating the energy source. I transformed these into single-level column names by concatenating the year and energy source (e.g., Solar_2022), ensuring that each column clearly represented both the year and the energy source.

Finally, I ensured that the power units were consistent across all columns and addressed any missing values to maintain data integrity.

3.1.2. Electricity generation by fuel

Firstly, to clean the dataset, I removed the unnecessary columns such as the aggregates, "Other" and "Other Renewables". These columns were irrelevant to the analysis and could cause issues moving forward. The missing values and placeholders (e.g., '^' indicating values less than 0.005) were replaced with zeroes to ensure consistency.

Secondly, rows with regional summaries and aggregates were removed because they were irrelevant to the analysis. Flattening the columns and headers was crucial to preparing the dataset to be analyzed. Ensuring the columns were standardized and consistent prevented potential issues from arising. Multiple columns shared the same name, so they were modified to state the fuel source followed by the year for consistency. For example, “Natural Gas” under the 2022 header was converted to “natural_gas_2022”.

3.1.3. Oil: Trade 2022 and 2023

Data cleaning for oil trade dataset began with eliminating duplicate content, extra descriptive wording or notes that weren’t critical for the analysis of the data. Once non-essential content was removed, I standardized the column names to make them more consistent. Every column was distinctly renamed based on the year and the trade variable it tracked, for example, crude imports, refined product imports, crude exports, or refined product exports. This made it easier to handle import of the dataset as well as to manipulate the data.

Upon completing column names adjustments, I went into dealing with number reporting formats across the inputs. This included addressing errors where crude and refined products were either put together or, in reverse, presented separately. I’ve also looked for gaps in the data, or values that were incomplete or missing and handled them as needed to avoid discrepancies in the analysis. When the data was clean and consistent, I stored the new dataset in its final version, ready for integration with other datasets or analysis.

3.1.4. Carbon: Carbon Dioxide Emissions from Energy

The dataset was first loaded into a pandas DataFrame, and irrelevant rows and columns were removed. Missing values were dropped using “dropna()”, and the last three columns and rows were discarded as they contained summary or unrelated data. The column headers were updated by setting the first row as the header and converting year columns to integers.

The cleaned dataset was then saved to a MySQL database using SQL “to_sql()” function, with the table being replaced each time the script is run. A SQL query was applied to filter out rows containing "Total," ensuring only relevant data remained. Finally, the cleaned data was exported as a CSV file for further use.

3.1.5. Renewable Energy - Solar and Wind Installed capacity by type

The data cleaning process for this dataset began by removing all extraneous information included in the Excel sheet, such as descriptive text explaining the meaning of each column. Once the unnecessary content was eliminated, I separated the two main tables (solar and wind) into individual Excel files, ensuring each table had its own dedicated file.

Next, I standardized the column names for both tables. This involved renaming each column to explicitly include the year and the type of capacity it referred to, ensuring clarity and avoiding aggregated columns. For instance, instead of generic or combined column names, each one was labeled distinctly to reflect the specific data it represented. Additionally, I renamed the first column in both tables to "country" to ensure consistency and compatibility for subsequent operations.

After completing these transformations, I saved each cleaned table in its respective Excel file. Using Python, I then uploaded both tables and merged them based on the "country" column, combining the data into a single, unified table. Finally, I saved the merged table, preparing it for upload into the database for further analysis or use.

3.2. Data exploration for each dataset (**for each member)**

3.2.1. Renewables - Generation by source

The Renewables - Generation by Source dataset provided insights into renewable energy production across multiple countries, highlighting trends and changes between 2022 and 2023. Renewable energy sources are categorised into wind, solar, hydro, and other renewables, with data measured in terawatt-hours (TWh).

One notable finding was the substantial increase in renewable energy production in Middle Eastern countries like Qatar, Saudi Arabia, and the UAE. This growth was primarily driven by significant investments in solar energy infrastructure, leveraging the region's abundant sunlight. For instance, Qatar has seen over 200% increase in renewable energy production in just one year which is mainly due to its large-scale solar projects.

Another key insight was Canada’s heavy reliance on hydroelectric power. In 2023, hydro accounted for 86% of Canada’s renewable energy production, a notable increase from 60% in 2021.

An interesting observation emerged when comparing renewable energy portfolios globally. Countries like the US, France, China, and India diversified their renewable energy production across wind, solar, hydro, and other renewables. In contrast, Middle Eastern countries and Canada showed a high dependency on specific sources solar and hydro, respectively. This diversity in renewable energy portfolios is essential for ensuring energy security, but it also reflects the influence of factors like geography, natural resources, policies, and technological progress.

Overall, this dataset provided valuable insights into renewable energy trends and challenges countries face when shifting to sustainable energy sources.

3.2.2. Electricity generation by fuel

This dataset focuses on electricity generation by fuel for 2022 and 2023. By analyzing this data, I can derive conclusions regarding trends, patterns and which fuel sources overpower electricity generation in certain regions.

- Query 1: What are the trends in fuel generation from 2022 to 2023?

We can observe the countries where electricity generation by fuel grew the most and least. Our key findings are that Germany had the most significant decrease of 25.20% in electricity generation from fuel from 2022 to 2023, leading the global initiative in climate change. Wind energy was Germany's largest source of electricity in 2023. France's total electricity generation grew by 9.47%, primarily nuclear. India was the second leading county in electricity generation growth with 6.15% given its large economy was expected.

- Query 2: Which countries dominate in electricity generation by fuel?

The most prominent players in the field are China, the US and India producing 7,723.6, 3,746.3 and 1,724.1 TWh respectively. With the growing economies and booming industries which rely on electricity, focusing on how the electricity is created is essential in policy making for future environmental change. Rapid industrialization and urbanization have led China to gain the top spot in electricity production by fuel. Increasing energy demands fueled China's adoption of fuel to generate electricity.

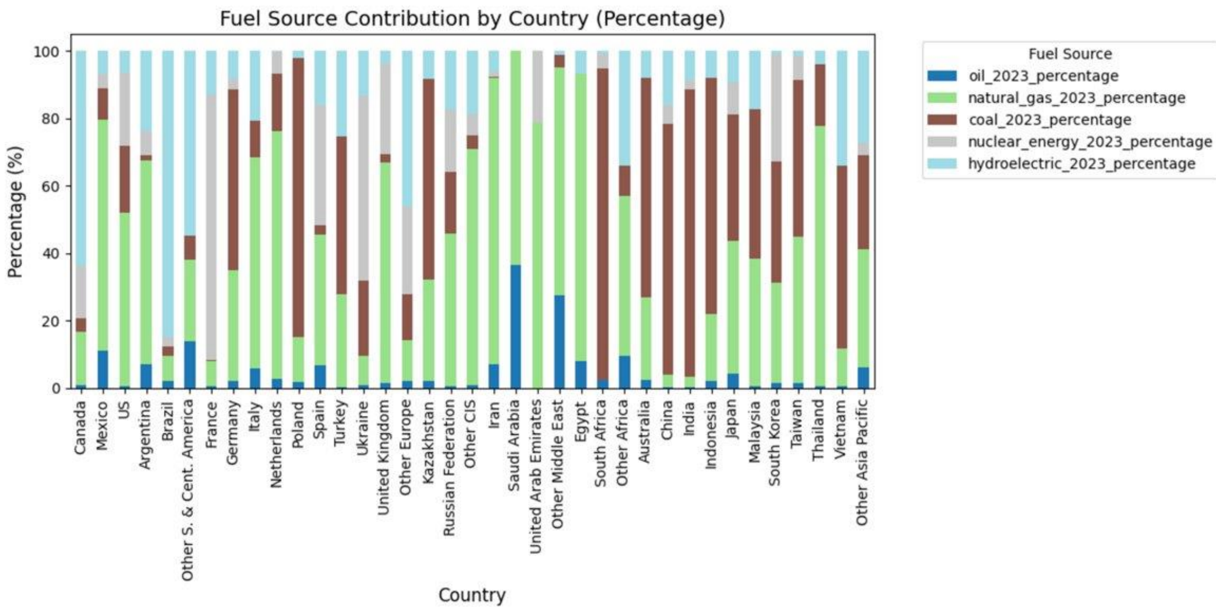
- Query 3: What is the dominant fuel type across countries for electricity generation?

Coal is the dominant fuel for electricity generation, with fossil fuels forming 60% of global electricity generation. Coal has been the cornerstone for power systems, particularly rapidly industrializing economies. Coal is found in abundance, is reliable and affordable. In addition, the infrastructure is often well-established.

- Query 4: Which country uses the most coal for electricity generation?

From the graph below it is evident that South Africa is leading in using coal to generate electricity with 92.35% of the total electricity being derived from coal. As shown in the table below, India, Poland and China also generated over 70% of their electricity from coal in 2023.

South Africa's reliance on coal has deep historical, economic and historical roots. The abundance of coal mines means that coal is cost-effective and the preferred energy source. The economy demands reliability in most sectors which in turn results in coal-intensive energy generation.



	country	total_generation_2023	coal_percent
0	South Africa	201.3	92.35
1	India	1724.1	85.34
2	Poland	121.0	82.81
3	China	7723.6	74.50

Energy and electricity are vital to our society and economy. Being able to analyze the trends will allow us to better understand how to transition to cleaner sources while keeping up with rising demands year over year.

3.2.3. Oil: Trade 2022 and 2023

Our project, titled "Global Trends in Energy and Emissions," focused on analyzing traditional energy sources such as crude oil and natural gas while comparing them to renewable energy sources to assess their impact on carbon emissions. The dataset I am working with provided detailed information on the movement of crude and refined oil products (imports and exports) for 2022 and 2023. By analyzing the mobility of these primary energy sources, we aimed to identify the key market players and correlate their activity to observed carbon emissions. In order to look and explore the data I've came up with five queries that helped me to dive deep into the understanding of the global oil trade.

- Query 1: Top Countries/Regions by Total Imports and Exports for Each Year.

First SQL query critical for the team was to understand top oil import and export (including both crude oil and petroleum refined products) countries between 2022 and 2023. Total petroleum imports and export volumes were calculated for each country and then sorted in descending order. That allowed to identify key players on the global energy market and also to identify trends for the subject years of analysis. This query also sets up the team for further analysis by providing a foundation of the top oil-trading nations. It was identified that United States, Saudi Arabia, Russian Federation, United Arab Emirates and Canada were top 5 crude and petroleum product exporters for both 2022 and 2023. On the other hand: Europe, China, Other Asia Pacific region, United States and India were among top 5 importers of petroleum goods in both 2022 and 2023.

- Query 2: Year-over-Year Growth in Imports and Exports for Each Country/Region.

Second query helped to expand the analysis of the available energy data by calculating annual change in crude and product imports and exports per country/region between 2022 and 2023. This query helped to identify the patterns that might relate to economical and geopolitical developments in the world. For instance, decline in imports might relate to reduction in refining capacity, or adversely relate to increase in renewables generation. We've observed that Iraq, Russian Federation, Kuwait, Saudi Arabia and China showed substantial increase in imports, while Singapore, East and South Africa, Other Asia Pacific, CIS and South and Central America regions showed year-on-year increase in exports within 2022-2023.

- Query 3: Countries/Regions with Highest Dependence on Product Imports and Exports.

Third query looked at the share of product imports and exports within each country's total oil trade (both crude and products) for 2022 and 2023. By pinpointing countries that rely heavily on product imports or exports, we get a clearer picture of their dependence on refined products rather than crude oil. Countries focusing more on products than crude are likely less involved in early-stage oil production but still play big roles in refining and downstream activities. This query helped to understand each country's place in the oil supply chain and opened up avenues to explore trade dependencies and energy policy impacts. Mexico, Iraq, West Africa, North Africa together with East and South Africa showed really high dependency on importing petroleum products revealing limitations on refining capabilities, while Japan, China, India, Singapore and Europe are among the top suppliers of the refined petroleum products (not crude).

- Query 4: Net Oil Trade Balance (Exports - Imports) for Each Country/Region.

Fourth query looked at each country's net trade balance, which we got by subtracting total imports from exports of both crude oil and refined products. Countries with a positive balance are exporting more than they're importing, while those with a negative balance are importing more than they export. Knowing who the net exporters and importers are helped to paint a picture of global oil trade and showed where different countries stand in terms of energy independence. This is useful for understanding trade relationships and dependencies between oil suppliers and those that rely heavily on imports. It has been identified that Saudi Arabia, Russian Federation, United Arab Emirates, Iraq and Canada were top net exporters of crude and refined petroleum products, while China, Europe, Other Asia Pacific region, India and Japan are highly dependent on imports of petroleum crude and products.

- Query 5: Top 5 Countries/Regions with the Largest Increase in Crude Exports from 2022 to 2023.

Last query related to oil trade dataset identified five countries that saw the biggest increase in crude oil exports from 2022 to 2023. Finding these top-growing exporters helped us see which countries might be putting more resources into oil production or focusing more on sending oil abroad. It's useful for spotting changes in the oil market and recognizing which countries are stepping up their presence. By highlighting these key exporters, we can take a closer look at why their exports have grown—whether due to new policies, expanded production facilities, or shifts in demand. South and Central America, United States, Other Asia Pacific region, Canada and Mexico saw the biggest increase in crude oil exports among the analyzed countries.

3.2.4. Carbon: Carbon Dioxide Emissions from Energy

The following queries were conducted to explore the carbon emissions dataset, providing insights into emission trends, regional comparisons, and the effects of global events on carbon output.

- Query 1: Impact of Covid-19 on Emissions (2015-2022)

This query analyzes carbon emissions from 2015 to 2022, focusing on the effects of the COVID-19 pandemic. By comparing the four years before the pandemic to the four years during it, this analysis highlights how the global slowdown in economic activity and mobility led to a temporary reduction in emissions. The results underscore the potential for achieving significant emission reductions during global disruptions, offering valuable insights into how human behavior and economic activity influence environmental outcomes.

- Query 2: Emissions in North America, Europe, and South America

This query compares carbon emissions across three regions: North America, Europe, and South America. By examining emissions in North America and Europe, we can assess how different energy policies and the adoption of renewable energy sources impact emission levels in these developed regions. The comparison with South America highlights how varying levels of industrialization and energy consumption patterns, particularly in terms of reliance on fossil fuels versus renewable sources, shape emissions in different parts of the world.

- Query 3: Ranking Countries by Total Emissions (1965-2023)

This query ranks countries based on their total carbon emissions from 1965 to 2023. The results reveal that the United States and China are the largest contributors to global emissions, with China's rapid industrialization leading to significant increases. This ranking provides a historical view of the major emitters and helps contextualize current discussions about climate change, highlighting the need for both developed and developing nations to work together in reducing global emissions.

- Query 4: Ranking Countries by Average Emissions (2019-2023)

This query ranks countries by their average carbon emissions over the five years from 2019 to 2023. The data shows that China, the United States, and India are the top emitters, reflecting ongoing industrialization and energy consumption trends. This analysis offers a contemporary view of global emissions, focusing on the countries that are currently making the largest contributions to climate change.

- Query 5: Emissions Changes Between 2022 and 2023

This query calculates the year-on-year change in carbon emissions from 2022 to 2023. The results reveal that countries like China and India have experienced significant increases in emissions, driven by continued industrial growth. In contrast, several developed nations, including the United States and Japan, have reduced their emissions, indicating progress toward more sustainable energy practices. This comparison provides valuable insight into how different countries are responding to global climate goals and transitioning toward lower-carbon economies.

3.2.5. Renewable Energy - Solar and Wind Installed capacity by type

To explore the dataset, I designed five specific queries aimed at gaining a deeper understanding of the data. Below, I outline the questions I sought to answer and the corresponding results:

- Query 1: Which countries have an increase in the installed photovoltaic power?

This query revealed that 53 countries experienced an increase in installed photovoltaic power between 2022 and 2023. This is a promising development, as it indicates that these countries are actively investing in solar energy infrastructure. Such investments pave the way for greater utilization of solar power in the future, contributing to cleaner and more sustainable energy generation. Among the countries identified are Canada, Mexico, US, Argentina, Brazil and Chile.

- Query 2: Which countries have a concentrated solar power in 2023 higher than the average?

This query revealed that nine countries have a concentrated solar power capacity exceeding the 2023 average. These nations are at the forefront of solar energy development, leading in solar capacity. As the project progresses, we aim to explore how this leadership correlates with other factors. The countries identified are US, Chile, Spain, Israel, United Arab Emirates, Morocco, South Africa, China and India.

- Query 3: Which countries have a decrease in the wind capacity onshore?

This query yielded an interesting observation: only one country, Uruguay, experienced a decrease in onshore wind capacity from 2022 to 2023. This finding is noteworthy because it indicates that all other countries have increased their onshore wind capacity during this period. It would be valuable to investigate the factors behind Uruguay's decline, exploring what occurred during these years and whether there is a specific explanation for this trend.

- Query 4: What is the top 5 countries with the most increase in wind capacity offshore?

This query aimed to identify the top five countries with the largest increase in offshore wind capacity. In the later stages of this project, we plan to investigate whether there is a relationship between this increase and changes in carbon emissions. The top five countries identified are China, Netherlands, Taiwan, United Kingdom and Denmark.

- Query 5: What is the Year-over-Year Growth Rate for Each Country for the installed photovoltaic power?

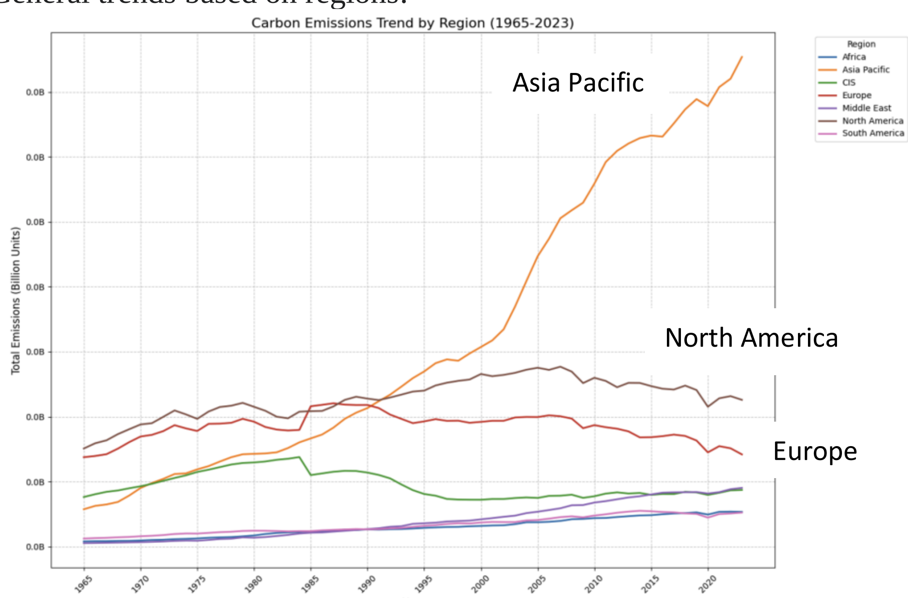
This query aimed to assess the percentage increase in installed photovoltaic power, providing a more comprehensive view of how much each country has expanded its capacity. We discovered that some countries have experienced remarkable increases, with values reaching as high as 473% for Saudi Arabia.

3.3. Answer guiding questions (**for each member**)

3.3.1. What are the general energy trends in carbon emissions?

To estimate the general energy trends in carbon emissions, the dataset was grouped by 7 regions instead of comparing by individual countries.

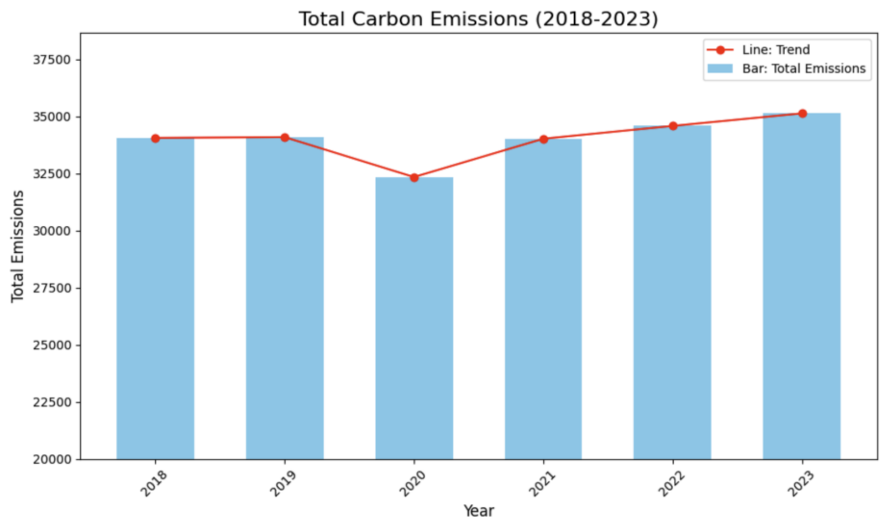
General trends based on regions:



One of the significant trends observed is the rapid growth of emissions in the Asia Pacific region. As shown in the line chart, this region has experienced the highest increase in emissions, driven primarily by the economic expansion and industrialization of countries like China and India (PricewaterhouseCoopers, n.d.).

North America and Europe follow in terms of emissions, but the gap between these regions and Asia Pacific is notable.

COVID-19 impact on carbon emissions:



Another key analysis focused on the impact of the COVID-19 pandemic on global emissions. The bar chart illustrating emissions from 2019 to 2020 reveals a significant drop in emissions during the pandemic, largely due to decreased economic activity and mobility (Liu et al., 2022). This temporary reduction highlights the potential for emission reductions during unprecedented global disruptions, providing a valuable lesson on how changes in human behavior and industrial activity can impact the environment.

Emission difference between 2022 and 2023:

```
"""
SELECT
    Country,
    ROUND(SUM(`2023`) - SUM(`2022`), 2) AS Emission_Difference
FROM carbon_emissions
GROUP BY Country
ORDER BY Emission_Difference DESC;
"""
```

	Country	Emission_Difference
0	US	-158.51
1	Japan	-68.38
2	Germany	-59.89
3	Poland	-24.25
4	Italy	-23.60

```
"""
SELECT
    Country,
    ROUND(SUM(`2023`) - SUM(`2022`), 2) AS Emission_Difference
FROM carbon_emissions
GROUP BY Country
ORDER BY Emission_Difference ASC;
"""
```

	Country	Emission_Difference
0	China	642.06
1	India	218.71
2	Vietnam	47.92
3	Mexico	20.60
4	Venezuela	16.18

Additionally, the analysis of emission changes between 2022 and 2023 identified countries with the highest increases and decreases in emissions. Notably, countries like China and India saw significant increases, continuing their pattern of rapid industrialization. In contrast, several developed nations, including the United States and Japan, showed reductions in their carbon footprints, indicating that these countries are successfully transitioning to cleaner energy sources and adopting more sustainable practices.

Conclusion:

Overall, these analyses reveal that while countries like China and India are still heavily reliant on industrial activity and fossil fuels, many developed nations are making strides in reducing their emissions. These insights are crucial for understanding how different regions and countries are responding to global environmental challenges and offer valuable lessons for policymakers aiming to combat climate change.

3.3.2. What is the relationship between renewable energy capacity and carbon emissions?

To answer this guiding question, we need to merge the carbon emissions table with the energy capacity table. Our focus will be on the changes between 2022 and 2023, as the energy capacity data only covers this time frame. Our approach will begin by comparing each energy capacity (solar and wind) individually with carbon emissions. Finally, we will combine the results to analyze the overall findings.

Solar capacity and carbon emissions:

The goal of this query is to examine the changes in carbon emissions for all countries that saw an increase in installed photovoltaic power from 2022 to 2023.

```
SELECT
    ec.country,
    (ec.solar_pv_2023 - ec.solar_pv_2022) AS increase_solar_pv,
    (ce.2023-ce.2022) AS increase_carbon_emission
FROM
    energy_capacity ec
JOIN
    carbon_emission ce ON ec.country = ce.country
WHERE
    ec.solar_pv_2023 > ec.solar_pv_2022
ORDER BY increase_carbon_emission ASC
```

The results show the top five countries with the greatest decrease in carbon emissions (on the left) and the top five countries with the largest increase in carbon emissions (on the right).

	country	increase_solar_pv	increase_carbon_emission
0	US	24844.330	-158.510143
1	Japan	4011.000	-68.382897
2	Germany	14260.000	-59.892917
3	Poland	3639.000	-24.252011
4	Italy	5234.000	-23.595783

42	Other Asia Pacific	971.196	12.224342
43	Russian Federation	354.220	15.678088
44	Mexico	1546.380	20.603600
45	India	9718.938	218.713226
46	China	216889.000	642.056029

It's a positive finding that 35 countries reduced their carbon emissions, while only 12 countries experienced an increase, and these are the countries that also saw growth in their installed photovoltaic power.

Wind capacity and carbon emissions:

Similar to the previous query, we aim to analyze the changes in carbon emissions from 2022 to 2023 for countries that experienced an increase in both onshore and offshore wind capacity.

```
SELECT
    ec.country,
    (ec.wind_onshore_2023 - ec.wind_onshore_2022) AS increase_wind_onshore,
    (ec.wind_offshore_2023 - ec.wind_offshore_2022) AS increase_wind_offshore,
    (ce.2023-ce.2022) AS increase_carbon_emission
FROM
    energy_capacity ec
JOIN
    carbon_emission ce ON ec.country = ce.country
WHERE
    ec.wind_onshore_2023 > ec.wind_onshore_2022
    AND ec.wind_offshore_2023 > ec.wind_offshore_2022
ORDER BY increase_carbon_emission ASC
```

The results show that only 9 countries have increased both onshore and offshore wind capacity. Of these, 2 countries have seen an increase in carbon emissions, while 7 countries have experienced a decrease in emissions.

	country	increase_wind_onshore	increase_wind_offshore	increase_carbon_emission
0	Japan	758.500	101.5	-68.382897
1	Germany	3038.000	258.0	-59.892917
2	United Kingdom	635.000	818.0	-13.821830
3	Spain	912.000	2.0	-13.511558
4	Netherlands	585.656	1408.5	-8.744101
5	Taiwan	268.767	824.0	-7.826198
6	Denmark	54.000	344.0	-1.333002
7	Other Asia Pacific	1060.250	110.0	12.224342
8	China	69100.760	6830.0	642.056029

Solar, wind and carbon emissions:

Finally, we will analyze the relationship between the increase in both solar and wind capacity and carbon emissions. For this query, we will merge the energy capacity and carbon emissions tables, calculate the increase in energy capacity, and determine the difference in carbon emissions between 2023 and 2022.

```sql
SELECT
    ec.country,
    (ec.solar_pv_2023 - ec.solar_pv_2022) AS increase_solar_pv,
    (ec.wind_onshore_2023 - ec.wind_onshore_2022) AS increase_wind_onshore,
    (ec.wind_offshore_2023 - ec.wind_offshore_2022) AS increase_wind_offshore,
    (ce.2023-ce.2022) AS increase_carbon_emission
FROM
    energy_capacity ec
JOIN
    carbon_emission ce ON ec.country = ce.country
WHERE
    ec.wind_onshore_2023 > ec.wind_onshore_2022
    AND ec.wind_offshore_2023 > ec.wind_offshore_2022
    AND ec.solar_pv_2023 > ec.solar_pv_2022
ORDER BY increase carbon emission ASC
```

Since the previous query yielded results for only 9 countries, we expect this query to include 9 or fewer countries as well. The results are as follows:

	country	increase_solar_pv	increase_wind_onshore	increase_wind_offshore	increase_carbon_emission
0	Japan	4011.000	758.500	101.5	-68.382897
1	Germany	14260.000	3038.000	258.0	-59.892917
2	United Kingdom	1005.500	635.000	818.0	-13.821830
3	Spain	5401.000	912.000	2.0	-13.511558
4	Netherlands	4304.193	585.656	1408.5	-8.744101
5	Taiwan	2693.938	268.767	824.0	-7.826198
6	Denmark	459.100	54.000	344.0	-1.333002
7	Other Asia Pacific	971.196	1060.250	110.0	12.224342
8	China	216889.000	69100.760	6830.0	642.056029

Conclusion:

Based on the results of these queries, we can conclude that there is a clear relationship between energy capacity and carbon emissions. We observed that, on average, 74% of the countries that increased their installed photovoltaic capacity also saw a reduction in carbon emissions. Similarly, approximately 77% of the countries that expanded their onshore and offshore wind capacity experienced a decrease in carbon emissions. These findings align with expectations and highlight how the transition to renewable energy sources can effectively reduce carbon emissions, thereby contributing to the mitigation of global warming.

3.3.3. How do the countries with the largest oil trade volumes contribute to global carbon emissions?

	country	total_imports_2022	total_exports_2022	carbon_emissions_2022
0	China	12143.40	1168.79	10576.32
1	US	8340.99	8548.56	4798.22
2	India	5752.94	1869.31	2595.61
3	Russian Federation	72.14	7821.13	1599.06
4	Japan	3465.16	348.73	1081.17
5	Saudi Arabia	200.31	8766.98	609.38
6	Canada	1140.10	4678.79	524.95
7	Mexico	1231.00	1173.10	469.27
8	United Arab Emirates	775.60	5467.33	281.38
9	Other Asia Pacific	9918.86	3559.06	218.24
10	Singapore	2401.77	1529.33	209.76

	country	total_imports_2023	total_exports_2023	carbon_emissions_2023
0	China	13716.53	1308.57	11218.37
1	US	8539.34	9108.46	4639.71
2	India	5769.78	1988.70	2814.32
3	Russian Federation	181.22	6735.54	1614.73
4	Japan	3332.17	260.74	1012.78
5	Saudi Arabia	187.63	8281.81	620.41
6	Canada	1073.40	4836.12	519.51
7	Mexico	1229.00	1271.29	489.87
8	United Arab Emirates	736.23	5362.62	287.17
9	Other Asia Pacific	9665.90	4010.63	230.46
10	Singapore	2358.09	1670.75	224.38

It is deemed that crude oil and refined petroleum products are the primary source of carbon emission. We wanted to see if that was the case if looking at the actual numbers. Based on the analysis performed utilizing oil trade and carbon emissions datasets we derived the following conclusions.

China dominates the list on emissions with the most crude and petroleum product imports (13,716 bbl/day) and very low exports (1,308 bbl/day). Such dependence on imports is related to the fact that it is the world's 2nd largest economy and 2nd most populated country with high demand on crude, what in turn leads to it being the biggest carbon dioxide emitter (11,218 Mt/year). Similarly, India has a high level of total imports (5,769.78 bbl/day) and relatively lower exports (1,988 bbl/day). As its carbon emissions (2,814 Mt/year) demonstrate, oil is one of its main contributors to its emissions, because of its growing industrialisation and population. The US stands alone as a top importer (8,539 bbl/day) and exporter (9,108 bbl/day), what suggests US having a very strong domestic production and trading market. Despite its significant exports as well as the efforts in renewable energy, it is still the second biggest carbon emitter (4,639 Mt/year), meaning that it heavily relies on oil for domestic consumption.

Saudi Arabia and the United Arab Emirates are dominant on exports (with 8,281 bbl/day and 5,362 bbl/day, respectively) while having low imports level. Their colossal oil exports leave them with relatively low domestic carbon emissions (620 Mt/year for Saudi Arabia and 287 Mt/year for UAE), suggesting that their presence as suppliers leads to indirect emissions in importing countries.
Russian Federation shares similar behaviour with high exports (6,735 bbl/day), low imports (181 bbl/day) and relatively low carbon emissions (1,614 million tonnes per year).

Canada, Singapore and Mexico exhibit a balance between imports and exports. Canada (1,073 bbl/day; 4,836 bbl/day), Singapore (2,358 bbl/day, 1,670 bbl/day) and Mexico (1,229 bbl/day; 1,271 bbl/day) have moderate carbon emissions (519, 224 and 489 Mt/year respectively).
This data clearly shows a distinction between oil importers with high levels of carbon emissions (China, India, US) and large oil exporters (Saudi Arabia, UAE, Russia) with lower levels of carbon emissions. This observation leads to a very important conclusion that emissions are primarily generated by importing countries consuming oil products and while exporting countries show low emission level, they do their share indirectly by supplying crude oil and refined products to the world market. In fact, there is 80-90% of the total lifecycle greenhouse gas emissions from crude oil are attributed to combustion (U.S. Environmental Protection Agency, n.d.).

3.3.4. Is there a noticeable impact on electricity generation fuel mixes in countries with significant renewable energy installations?

This analysis reveals that countries with significant renewable energy installations are experiencing a noticeable impact on their electricity generation fuel mixes. While some nations have reduced their reliance on fossil fuels, others still heavily depend on traditional sources like coal and natural gas, despite substantial investments in renewables.

- **Middle Eastern Countries**

Middle Eastern countries, such as Qatar, Saudi Arabia, and the UAE, have made significant investments in solar energy which paved the path for a substantial increase in renewable energy production.

- **Qatar:** Renewable energy production increased by 250% from 2022 to 2023.
- **Saudi Arabia:** Renewables grew by 152.9%, primarily due to solar energy investments.

Even though these countries have seen a surge in renewable energy production, fossil fuels continue to dominate their energy mixes:

- **Saudi Arabia:** Fossil fuels accounted for 98.6% of its energy mix in 2023, with 62.7% coming from natural gas and 36% from oil.
- **UAE:** Fossil fuel reliance was reduced by 5% in 2023, a step in the right direction one might say. However, 72.1% of its energy mix still comes from fossil fuels. The UAE has diversified its energy mix, with 19.6% of its energy now derived from nuclear power.

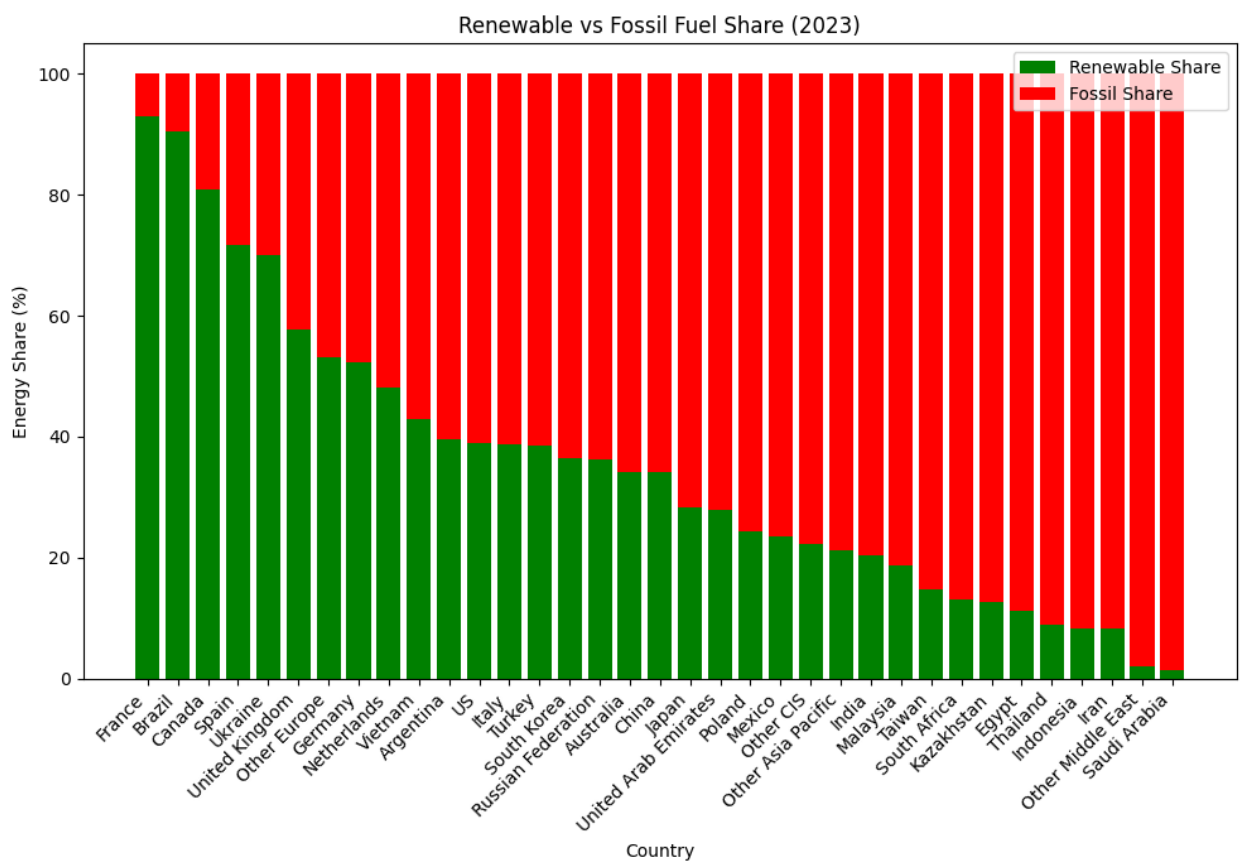
These findings suggest that while Middle Eastern countries are making progress in renewable energy, transitioning to a cleaner energy mix remains a challenge.

- **European Countries**

The data shows that European countries have made significant strides in transitioning to renewable energy sources, leading to reductions in fossil fuel usage.

- France: Renewable energy production increased by 25.9% in 2023, driven primarily by wind, hydro, and nuclear power. Nuclear energy accounted for nearly 70% of France's energy mix, and fossil fuel usage was reduced by 31.4%.
- Spain: Renewable energy grew by 15.6%, with wind and solar as major contributors. Spain has a diversified energy mix, with 22% from natural gas, 23% from wind, 17% from solar, and 21% from nuclear power.
- Germany: Renewable energy increased by 17%, while fossil fuel usage decreased by 18.9%, reflecting Germany's commitment to energy transition. In 2023, Germany's energy mix was almost even with 52% renewable/clean energy sources and 48% fossil fuels.

These findings highlight Europe's leadership in reducing its reliance on fossil fuels and increasing renewable energy production.

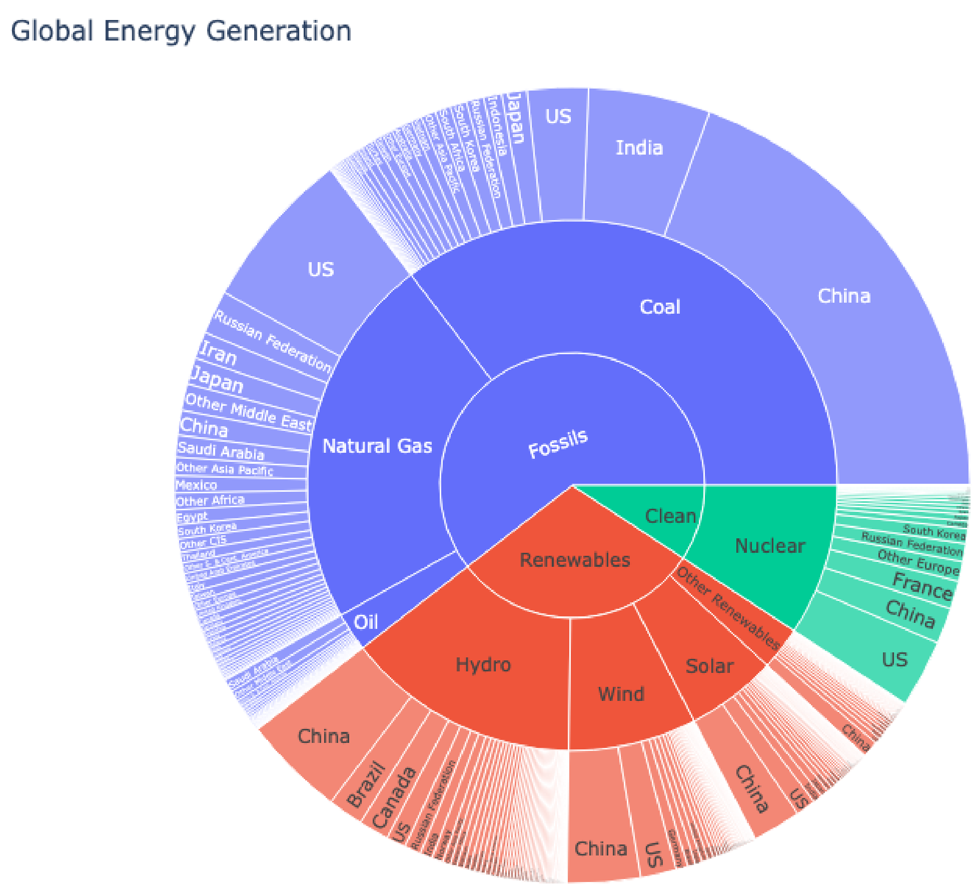


- **Global Energy Giants**

Global energy giants, such as China and the United States, exhibit high energy production levels but face challenges in reducing reliance on fossil fuels.

- **China:** In 2023, China generated 9193.6 TWh of electricity, to put this in perspective it is more than the combined output of all European nations in the year 2023. It also leads in renewable energy production, generating 2894.6 TWh in 2023. While this is a significant number, 65% of its energy still comes from fossil fuels, underscoring the challenges of transitioning to cleaner sources in high-demand countries.
- **United States:** The US follow a similar trend as China producing 973.7 TWh of renewable energy in 2023, a figure similar to the previous year. Fossil fuels still dominate its energy mix, with 43% from natural gas and 20% from nuclear power.

It's always interesting how just looking at the numbers can sometimes be misleading. While China and the US are the top renewable energy producers, they still have a long way to go in transitioning to cleaner energy sources. This analysis highlights the importance of not only increasing renewable energy production but also reducing reliance on fossil fuels to achieve a sustainable energy future.



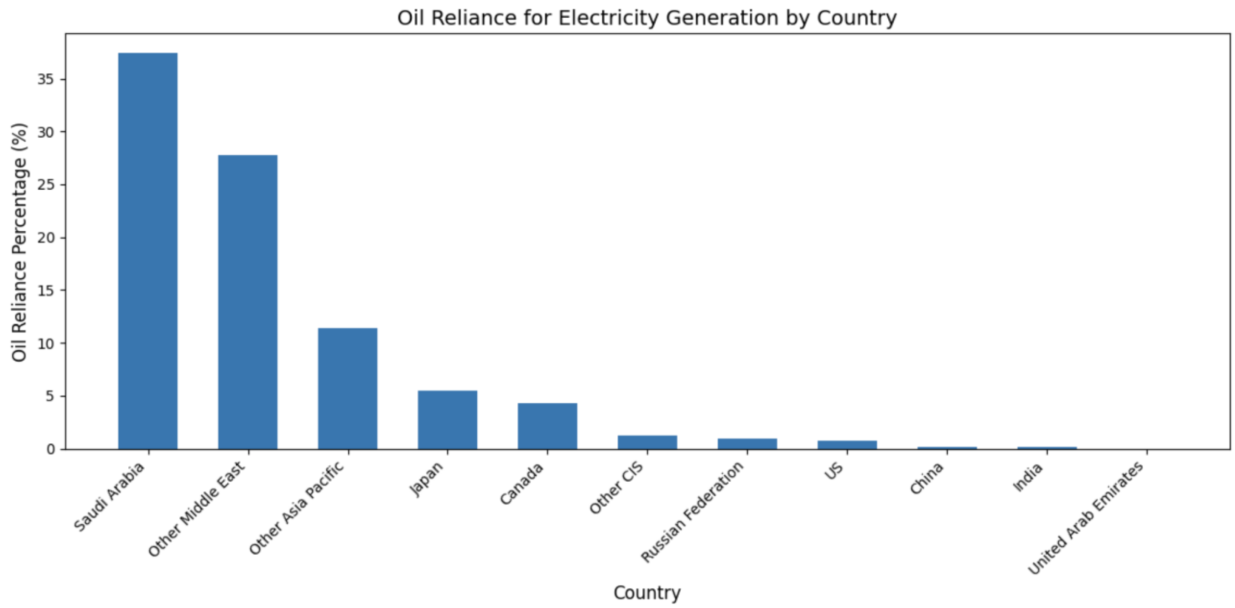
The global green energy landscape is evolving, with countries making significant strides in renewable energy production. However, there is still a long way to go in achieving a sustainable energy future. International cooperation and new policies are essential to accelerate this transition.

3.3.5. How do countries heavily involved in oil imports or exports compare in their reliance on oil for electricity generation?

Saudi Arabia and Other Middle Eastern countries lead in oil reliance for electricity generation. Saudi Arabia tops the chart is 37.4%. These countries have abundance of oil reserves which make oil a cost-effective source for electricity generation. Even though these countries are large players in the oil imports and exports, the lack of other substantial sources such as hydropower led to oil being a vital source in their energy generation.

China and India have minimal reliance on oil, 0.193% and 0.177% respectively. They are large importers of oil, but the electricity generation does not rely on oil. These countries from my analysis prior have shown to rely heavily on coal and renewables to meet their energy demands. Canada and the US are not oil reliant for their electricity generation having been major contenders in the oil import and exports.

There are economic incentives which support the trends. Saudi Arabia's reliance on oil persists as the infrastructure lacks diversification. Using oil is cheaper than investing in other forms to diversify electricity generation. On the other, large importers such as China aren't reliant on oil for electricity generation as it would be unstable due to high costs of importing and lack of stability.



	country	oil_electricity_2022	oil_electricity_2023	oil_reliance_percentage	Crude_Imports_2022	Crude_Exports_2022	Product_Import_2022	Product_Exports_2022
0	Saudi Arabia	153.0	152.1	37.385124	34.706363	7386.352942	165.606042	1380.622438
1	Other Middle East	124.7	129.9	27.761422	303.927021	1873.935716	526.924302	1295.724479
2	Other Asia Pacific	60.4	39.5	11.405412	5977.037368	987.810264	3941.823331	2571.248788
3	Japan	42.5	33.4	5.507183	2661.120155	8.562446	804.038426	340.164039
4	Canada	5.2	4.8	4.266212	518.436662	3993.939893	621.666222	684.847905
5	Other CIS	2.3	2.1	1.269109	324.686375	1811.892951	159.761040	257.446126
6	Russian Federation	7.0	6.8	0.933063	36.593714	5274.826429	35.547883	2546.305174
7	US	24.7	17.7	0.779512	6281.296137	3240.182047	2059.693116	5308.373032
8	China	11.5	11.2	0.192983	10206.423550	33.293121	1936.975195	1135.498242
9	India	2.4	2.8	0.177487	4642.304460	43.428582	1110.639612	1825.883502
10	United Arab Emirates	0.0	0.0	0.000000	150.492462	3583.574249	625.104902	1883.751109

4. **Discussion** (**for each member**)

4.1. What we learned

4.1.1. Pawan Bath

By analyzing and working with databases during this term, I’ve learned the importance of properly cleaning and understanding datasets before diving deeper into analysis. I’ve gained a deeper understanding of database structure, including the importance of tables and primary keys. Being able to join datasets has been a powerful skill I will implement in my future endeavours.

Problem solving is a vital skill that I’ve been able to hone during this project. Breaking big problems into smaller, manageable sections has been invaluable. Debugging is another branch which allowed me to practice what I had learnt in class.

Gaining hands on experience writing queries with SQL, filtering, sorting and aggregating data has been crucial to deriving insights which I can be confident in. The importance and power of SQL was validated as I performed queries to derive insights and explore my dataset. My technical skills have been refined alongside my teamworking skills.

4.1.2. Vladimir Stepanov

Working with databases during this project highlighted the importance of fully understanding the structure and connections within the data before diving into analysis. By identifying "country" as the common column across tables, we were able to flawlessly join datasets, which made it easier to analyze oil trade volumes together with carbon emissions.

I also gained hands-on experience with SQL to aggregate, filter, and sort data, what allowed to uncover trends in oil trade and emissions. SQL combined with Python helped to streamline the process, from querying to visualizing results. This reinforced the value of databases for handling several datasets while allowing to extract meaningful information. It showed me how powerful SQL is for analyzing the data and performing thorough investigation on available material.

4.1.3. Yuhan Zhang

In analyzing this dataset, I gained valuable experience with data cleansing, conversion and querying. Transforming the dataset into a structured SQL table required dealing with missing values, reshaping the data for better usability and ensuring that the dataset was prepared for accurate analysis. Developing specific queries, e.g. on regional trends, rankings and the impact of global events such as COVID-19, provided a deeper understanding of how to gain meaningful insights. In addition, visualizing trends using charts and interpreting the results required refining skills to present complex data in a clear and actionable way. This project emphasized the importance of careful data preparation and thoughtful query design to uncover trends and draw meaningful conclusions from large data sets.

4.1.4. Vhanessa Cardona

I learned that when working with databases, it's crucial to first thoroughly understand the data in all your tables and derive insights from them. This foundational understanding allows you to formulate meaningful questions that can be answered by querying other tables within the same database. Additionally, I gained experience using Python to access a database, which was a new skill for me. I also discovered that once you have all your tables, it’s important to understand how they are interconnected and how you can establish relationships between them. In this project, this was straightforward because all of our tables shared a common column (country). Ultimately, I realized that working with databases is an efficient way to retrieve and analyze data, as it provides a seamless way to query and answer critical questions.

4.1.5. Malhar Dhawle

Through this project, I gained valuable experience in using SQL and Python. I also learned how various methods through which I can create and connect to a SQL database using Python. I also learned how to clean and preprocess data especially when normalizing column names and handling missing values. Leveraging SQL join function across multiple tables to get a deeper understanding of the data.

In my opinion working with this database highlighted how misleading just looking at the numbers can be and how important it is to have a broad understanding of the data. If had just looked at Renewable energy production I would have thought Middle Eastern countries and China, or the US are taking significant steps in renewable energy production. However, when I looked at the other tables in our database, I realized that they still have a long way to go in transitioning to cleaner energy sources. This project has given me a better understanding of the global energy landscape and the challenges in transitioning to cleaner energy.

4.2. What we could have done differently

4.2.1. Pawan Bath

Looking back on the last six weeks, there are a few improvements we could have made. There were various paths we could have chosen as there were many different tables in the file. We could have spent more time looking into other datasets such as economic, solar and wind. This would have allowed us to come up with more interesting guiding questions.

4.2.2. Vladimir Stepanov

While looking back, there are a few areas that we could've done better. First, we could've spent more time investigating data cleansing and preprocessing methods to maintain consistency across tables. The slightest errors – unit differences or different column names – sometimes interfered with the analysis and could have been handled more efficiently when they first emerged. Standardising the data first would have greatly facilitated future queries and analysis. Also, a closer look at other datasets (wind energy, or economics, for instance) would've helped to give more clarity on how oil trade influences global carbon emissions.

4.2.3. Yuhan Zhang

Reflecting on the project, one improvement could have been the inclusion of additional economic data to enhance the analysis. For instance, incorporating GDP, energy prices, or industrial production data could provide context to the emissions trends observed in different regions. These variables would help establish a clearer relationship between economic activity and carbon emissions, offering deeper insights into how economic growth or downturns influence emission levels.

4.2.4. Vhanessa Cardona

One aspect we could have approached differently is the selection of the dataset. We discovered a dataset in an Excel file that contained multiple tables. While we chose to focus our analysis on carbon emissions and energy generation, there were other tables that could have led to alternative analyses. Regarding the dataset selection process, we aimed to include tables with diverse columns, which ultimately guided our decision on which tables to retain for analysis.

4.2.5. Malhar Dhawle

Reflection on the project, I think there are a few areas I am still interested in exploring. One area, which a few of my peers also mentioned, is examining how Gross Domestic Product (GDP) influences a country's energy generation mix and carbon emissions. This interest in GDP also stems from a previous course, DATA 601, where we analyzed the relationship between a country's GDP and their performance in the Olympics. Though these topics are very different, it would be insightful to explore whether higher GDP correlates with faster adoption of renewables.

Another area would be including geographical data to highlight how location shapes energy strategies. Finally, I am curious about how government policies and international tensions influence a country's energy mix. While I understand that acquiring and analyzing government policies and international tensions data would be challenging, given that it varies by country and changes constantly, it would still be a fascinating area to explore.

4.3. Potential opportunities for future work

4.3.1. Pawan Bath

Moving forward, there are various tables in the file which can be analyzed to extend our work. One suggestion could be to use the tables which describe electricity generation by oil, gas, coal and other from 1985. This analysis will allow us to gain a deeper understanding of patterns and trends over time.

4.3.2. Vladimir Stepanov

In the future, there's room to refine the analysis using more datasets — complete renewable energy use, economic growth, population growth. It would give us a greater understanding of how oil trade, carbon emissions and sustainable development relate to each other. Geospatial data would also be helpful to show trade paths and emission hotspots via properly selected diagrams. Data cleaning and collection could be automated, thus making it easier to analyze data in real time and see the results much faster.

4.3.3. Yuhan Zhang

One potential opportunity could be integrating policy and economic data, such as carbon taxes, energy subsidies, or economic growth rates, to assess how these factors shape emissions patterns. Analyzing the effectiveness of energy policies in reducing emissions could provide actionable insights for governments and organizations.

4.3.4. Vhanessa Cardona

As mentioned earlier, the file contains multiple tables, and for future work, we could explore these additional tables to broaden the scope of our project. One potential direction is to conduct a time-series analysis, as the data spans from 1965 to 2023. Additionally, we could apply machine learning techniques to generate predictions based on this data.

4.3.5. Malhar Dhawle

An area of potential future work I am interested is using cloud services for the storage and management of our database. I am particularly interested in leveraging AWS, specifically services such as Amazon Relational Database Service (RDS) and Simple Storage Service (S3) for data storage. Additionally, I would like to explore how we can work with the data stored in these cloud services. Finally, I aim to use the built-in analytics tools (AWS Quicksight) and machine learning services (AWS Sagemaker) available to gain deeper insights into our data. This project would be a valuable learning experience, preparing us for real-world projects.

5. Conclusion

5.1. Outcome of the project

Our main goal for this project was to explore the relationships between renewable energy, oil trade, electricity generation and carbon emissions. Combined insights derived from each individual topic allowed the team to identify important patterns and correlations that helped to grasp an understanding on global energy sector.

While looking at electricity generation by fuel and petroleum trade data we've identified that coal, natural gas and crude oil still continue to be the primary sources of power generation in many countries. Low cost, wide availability and existing infrastructure for conventional fuels provide little to no incentive for developing countries to invest in alternative energy sources. Based on the observed numbers even developed regions like US and Europe, that lead the transition to cleaner energy, continue to be the highest importers of crude and petroleum products contributing significantly to overall emissions.

Nonetheless, solar and wind capacity numbers revealed promising trends. Between 2022 and 2023, nearly three-quarters of countries that expanded their solar installations, and over two-thirds that grew their wind capacity, saw reductions in carbon emissions showcasing stories of success in trying to meet net zero targets.

By and large, while there is obvious progress and growth in renewable energy solutions across the world, there are still significant challenges and bottlenecks that exist in meeting "Paris Accord" commitments. Bloomberg estimated that over a 3 trillion dollar of additional investments are required annually in order to achieve the set goals by 2050 (Bloomberg, 2024). While complete phase out of traditional fuels might not be realistic, continued investment in renewable energy supported by stronger international policies and global cooperation is essential for bringing renewable sources in the forefront of the energy supply, thus, building a better future for next generations.

5.2. Outcomes of the data exploration (Guiding question)

Exploring individual datasets revealed insights into renewable energy production, carbon emissions, and oil trade trends. Middle Eastern countries showed rapid growth in renewable energy production, driven by substantial investments in solar energy infrastructure. Other countries like the US, Spain, South Africa, China, and India with highly concentrated solar power capacities also showcased advancements in solar energy, reflecting a global shift toward cleaner energy sources.

In the oil trade, the United States, Saudi Arabia, the Russian Federation, the UAE, and Canada were identified as the top five exporters of crude. On the other hand, Europe, China, the Other Asia Pacific region, and India ranked as the top importers. South and Central America had a significant growth in crude exports which can be due to` various factors like geopolitical tensions, economic growth, and technological advancements. We identified countries like Mexico, Iraq, and West Africa that showed high dependency on refined petroleum products, reflecting on refining capabilities.

Our exploration also revealed contrasts in global energy trends. While some countries are making significant strides in renewable energy production, others still heavily rely on traditional fossil fuels. We found a strong relationship between renewable energy capacity and carbon emissions. For instance, solar energy capacity showed an inverse relationship with emissions in 74% of the countries, except for those experiencing rapid industrialization and ongoing international tensions. Where nations like China and India with their fast-industrializing experienced significant increases in emissions, while developed countries such as the United States and Japan showed progress in reducing emissions. A similar pattern was observed between wind capacity and carbon emissions. Oil trade analysis highlighted how high exporters like Saudi Arabia and Russia have relatively low domestic carbon emissions but contribute indirectly to emissions in importing countries. Also, a temporary decline in global emissions was recorded due to reduced economic activity during the COVID-19 pandemic.

The insights from this study highlight the need for global cooperation and reveal disparities in energy practices across countries. It also serves as a reminder of the urgent need to address climate change and promote sustainable energy practices.

5.3. What the team learned in the process of completing the project

Through the course of completing this project, we gained valuable insights into the structure and functionality of relational databases as well as the practical use of SQL for data retrieval and analysis. Specifically, we explored how to query information from individual tables and how to join multiple tables to derive more complex insights. One key takeaway about databases was the importance of having a common key to link tables. In our case, all the tables shared a common column, country, which served as the key for our joins.

Additionally, we delved into the topic of Global Trends in Power and Emissions, uncovering several significant insights. For instance, we observed a global reduction in carbon emissions during the COVID-19 pandemic, reflecting the widespread slowdown in industrial and transportation activities. We also noted that many countries are actively transitioning to renewable energy sources. Furthermore, our analysis revealed key patterns in oil trade: the United States emerged as a unique case, standing out as both a leading importer and exporter of oil, indicating a robust domestic production and trading market. We also identified the top oil-importing and exporting nations globally and discovered that the Asia-Pacific region has seen a notable rise in carbon emissions compared to other regions.

6. References

- Energy Institute. (n.d.). *Statistical Review of World Energy: Resources and Data Downloads*. Retrieved October 24, 2024 from https://www.energyinst.org/statistical-review/resources-and-data-downloads
- Environment, U. (n.d.). *Energy*. UNEP. https://www.unep.org/topics/energy
- Executive Summary – *World Energy Outlook 2024 – Analysis - IEA*. (n.d.). IEA. https://www.iea.org/reports/world-energy-outlook-2024/executive-summary
- Liu, Z., Deng, Z., Zhu, B., Ciais, P., Davis, S. J., Tan, J., Andrew, R. M., Boucher, O., Arous, S. B., Canadell, J. G., Dou, X., Friedlingstein, P., Gentine, P., Guo, R., Hong, C., Jackson, R. B., Kammen, D. M., Ke, P., Quéré, C. L., . . . Schellnhuber, H. J. (2022). Global patterns of daily CO2 emissions reductions in the first year of COVID-19. *Nature Geoscience*, *15*(8), 615–620. https://doi.org/10.1038/s41561-022-00965-8
- PricewaterhouseCoopers. (n.d.). *The PWC Net Zero Economy Index 2023 – Asia Pacific cut*. PwC. https://www.pwc.com/gx/en/issues/esg/esg-asia-pacific/net-zero-economy-index-asia-pacifics-transition-2023.html
- Tierney, S. (n.d.). *Setting the record straight about renewable energy*. World Resources Institute. https://www.wri.org/insights/setting-record-straight-about-renewable-energy
- Bloomberg. (2020, June 19). A $3 trillion plan to hit Paris Accord goals: Week in Green. Bloomberg News. https://news.bloomberglaw.com/environment-and-energy/a-3-trillion-plan-to-hit-paris-accord-goals-week-in-green
- Oğuz, S. (2024, March 12). The $3 trillion clean energy investment gap, visualized. Visual Capitalist. https://decarbonization.visualcapitalist.com/the-3-trillion-clean-energy-investment-gap-visualized/
- U.S. Environmental Protection Agency. (n.d.). Lifecycle Greenhouse Gas Results. Retrieved from https://www.epa.gov/fuels-registration-reporting-and-compliance-help/lifecycle-greenhouse-gas-results