EXPLORATORY DATA ANALYSIS PROJECT REPORT

TITLE: EXPLORING THE DATA OF ENERGY CONSUMPTION/PRODUCTION, CO2 EMISSION AROUND THE WORLD (1980-2020)

SUBMITTED BY:

NAME: SHAIK ABDUL RASHEED

COURSE CODE: INT353

SECTION: K21UP

ROLL NO: 59

REG NO: 12111371

LOVELY PROFESSIONAL UNIVERSITY
Transforming Education Transforming India

Table of Contents

INTRODUCTION

Dataset: Energy Consumption/Production, Co2 emission around the world

Exploratory Data Analysis (EDA) is a crucial phase in understanding and interpreting complex datasets. In the context of energy consumption, production, and CO2 emissions around the world, EDA serves as a powerful tool to uncover patterns, trends, and insights. This introduction outlines the key aspects of the dataset and provides a context for the EDA project. The dataset under consideration for this EDA project encompasses global information on energy consumption, production, and CO2 emissions. It likely includes data from various countries, regions, or continents and spans a specific timeframe. The variables within the dataset may cover diverse aspects such as energy sources, economic indicators, population, and environmental metrics The primary goal is to explore trends in energy consumption, production, and CO2 emissions over time and across different geographical entities. This involves identifying patterns and variations that can inform our understanding of how energy dynamics impact the world. The EDA project aims to uncover regional variations in energy-related metrics. Examining disparities in consumption patterns, production sources, and emission levels between countries or continents can shed light on the global distribution of energy resources and environmental impacts Exploring relationships between variables within the dataset, such as the correlation between economic indicators and energy consumption, provides insights into the interconnected nature of these factors. Understanding these correlations is crucial for designing effective policies and interventions. Given the global emphasis on sustainability, the EDA project may delve into the adoption and growth of renewable energy sources. Analyzing the trends in renewable energy production and its impact on CO2 emissions can provide valuable insights into the progress toward cleaner energy.

DOMAIN KNOWLEDGE

Exploring a dataset on global energy consumption, production, and CO2 emissions necessitates a foundational understanding of several key domains. Fossil fuels, encompassing coal, oil, and natural gas, are primary contributors to energy production and carbon emissions, while renewables like solar and wind play an increasingly vital role. Recognizing the correlation between a country's GDP and energy demand is essential, as is considering the influence of population growth on consumption trends. Understanding the implications of CO2 emissions on climate change and the international initiatives addressing this issue, such as the Paris Agreement, provides context. Regional disparities in consumption patterns between developed and developing nations highlight the role of infrastructure and lifestyle differences. Accurate data sourcing from reputable agencies, awareness of potential inaccuracies, and an understanding of the influence of policies and technological advancements are critical. Exploring historical trends and recent shifts, especially in the context of a global transition to renewable energy, adds depth to the analysis. In essence, a multidimensional understanding of these domains enriches the Exploratory Data Analysis, offering nuanced insights into the intricate dynamics of global energy landscapes. Global energy consumption, production, and CO2 emissions have all increased steadily from 1980 to 2020. In 1980, global energy consumption was 72,500 TWh, global energy production was 67,700 TWh, and global CO2 emissions from energy consumption were 19.1 billion tonnes. By 2020, these figures had increased to 156,700 TWh, 150,700 TWh, and 36.3 billion tonnes, respectively. Exploratory data analysis (EDA) can be used to gain insights into the trends and relationships in the data on energy consumption/production, CO2 emissions, and other energy-related variables. Some examples of EDA that can be performed on this data include Visualize the data to identify patterns and trends, correlate different variables to identify relationships between them. EDA is a powerful tool for gaining insights into energy consumption/production, CO2 emissions, and other energy-related data.

Data Description:

The Dataset on Energy Consumption/ Production, Co2 emission have total of 55440 rows and 11 columns

1.Sr. No: Gives each row a unique number of each country with different energy types for different years

2.Country: Represents each country around the world at the time of 1980-2020

3.Energy_type: This shows each different types of that are there in nature and used for getting overall data on energy consumption and production

4.Year: This used as data for each year for countries energy every single time

5.Energy consumption: This column provides information on amount of energy consumed by each country and overall world

6.Energy production: This column gives us information on how much of energy of each type is being produced for utilizing it

7.GDP: the GDP (Gross Domestic Product) column is a crucial economic indicator providing insights into the economic performance of different countries or regions. GDP represents the total market value of all goods and services produced within a country over a specific time period.

8.Population: Population data is instrumental in understanding the scale and scope of energy consumption and CO2 emissions in relation to the size of a given population.

9.Energy intensity per capita: The "Energy Intensity per Capita" column in the dataset is a calculated metric that provides a measure of the energy efficiency of a country or region in relation to its population size.

10.Energy intensity by GDP: The "Energy Intensity by GDP" column in the dataset is a calculated metric that provides a measure of the energy efficiency of a country

11.CO2 emission: The "CO2 Emission" column in the dataset represents the amount of carbon dioxide (CO2) released into the atmosphere due to various human activities, such as energy production, industrial processes

Reason to select this Dataset:

Selecting the Energy Consumption/Production and CO2 Emission dataset spanning from 1980 to 2020 holds significant merit for various reasons. Firstly, the dataset's global scope provides a comprehensive view of energy dynamics and CO2 emissions, allowing for cross-country comparisons and insights into worldwide trends. The longitudinal nature of the dataset, covering over four decades, facilitates the identification of patterns and shifts in energy consumption and production, offering a valuable historical context. Analyzing CO2 emissions over this timeframe contributes to a nuanced understanding of the environmental impact of energy-related activities and provides insights into the effectiveness of global efforts to mitigate climate change. Moreover, the dataset likely captures the transformative period marked by the shift from conventional fossil fuel-based energy systems to more sustainable alternatives, enabling an examination of the progress in renewable energy adoption. By incorporating variables such as GDP and population, the dataset allows for a holistic exploration of the interconnections between economic development, demographic trends, and energy dynamics. This dataset, therefore, not only offers valuable insights into global energy landscapes but also serves as an educational resource and a foundation for informed discussions on sustainable development and environmental policies We will also explore correlations between energy consumption and economic indicators, providing insights into the interplay between economic growth and energy demand. Furthermore, our analysis will consider the environmental implications by examining associated greenhouse gas emissions. Ultimately, the EDA process will yield valuable insights that inform energy policy decisions, resource allocation, and sustainability initiatives, contributing to a more informed and responsible approach to energy consumption management.

Libraries used and Approaches

•**PANDAS**: Pandas is a powerful data manipulation library that allows you to load, clean, and manipulate data efficiently. It's often used to create data frames, filter, sort, and summarize data. I have used pandas library to read my csv file and to find head tail what are my columns and rows of csv file and drop some of the useless columns which are present there.

•**MATPLOTLIB**: Matplotlib is a popular data visualization library in Python that provides a wide range of plotting options, from basic line charts to complex heatmaps. I have used pie chart to show the type of content available and country, bar graphs to show genre, ratings

•**SEABORN**: Seaborn is built on top of Matplotlib and offers a high-level interface for creating informative and attractive statistical graphics. It simplifies the creation of aesthetically pleasing visualizations. It offers a higher-level interface for creating aesthetically pleasing and informative statistical graphics. We are using this library to enhance and to make more advanced graph like heat maps to improve our visualization

•**PLOTLY.EXPRESS**:Plotly Express is a high-level interface for creating interactive visualizations. It is easy to use and allows you to create various types of plots with minimal code.

•**PLOTLY.SUBPLOT**: Plotly Subplots is a part of the Plotly library and is used to create subplots within a single figure. It's especially useful when you want to display multiple plots in a grid layout

•**MATH**: math is a built-in module in the Python 3 standard library that provides standard mathematical constants and functions. You can use the math module to perform various mathematical calculations, such as numeric, trigonometric, logarithmic, and exponential calculations

•**SCIPY.STATS**: This module contains a large number of probability distributions, summary and frequency statistics, correlation functions and statistical tests, masked statistics, kernel density estimation, quasi-Monte Carlo functionality, and more.

•**IPYTHON.DISPLAY**: It is an interactive command-line terminal for Python

Steps of EDA

1. Data Collection and Loading:

df is a pandas DataFrame created to store data from “energy[1].csv”

2.Data Inspection:

Examine Basic information about the dataset df.shape function gives output of no of column and rows of the data set,df.describe() describes bout the columns in dataset,df.info() gives us information on dataset

3.Data column Renaming:

Renaming columns for simplicity Some column names are quite long so we are going to shorten the names of it and also renaming the energy type names

4.Converting Year to datetime:

Converting the year data in date time for more simplicity

5.Data Cleaning:

Firstly we will find for null values by giving isnull function and for cleaning this we will use dropna or fillna function to remove/fill null values like NaN in the columns

Now we will look duplicate values in each column and remove the after doing all this we will again recheck for null,duplicate values because if they are present in data set they may not give us appropriate output if there are no duplicate,null values we will continue for analyzing the dataset

6.Data Analysis:

a) Examining the "Year" column enables the identification of temporal trends in energy consumption, production, and CO2 emissions. This analysis can reveal patterns of growth, decline, or fluctuations over time, providing a historical context to the dataset

b)Analysing the energy type provides us critical lens for understanding the composition and evolution of the global energy landscape. Analyzing this variable offers insights into sustainability, environmental impact, technological advancements, and the effectiveness of energy policies at various scales.

c)Energy consumption may tells us the top countries that are consuming tha all energy which indicates that big countries like china’s consumption of one energy is equal to total energy consumtion of energy of a single small countries.

d) the "Energy Production" column is a key variable that enables a multifaceted analysis of energy dynamics. From understanding the total energy output to evaluating the composition of the energy mix and assessing the impact of policies, this variable provides critical insights into the sustainability and efficiency of a region's energy infrastructure.

e) The "GDP" (Gross Domestic Product) column in the dataset represents the economic output of a country or region. Analyzing this column can provide valuable insights into the intersection of economic development and energy dynamics.

f)By analysing population column it provides insights into the complex interactions between population dynamics, economic development, and energy consumption. It is a critical factor in understanding the challenges and opportunities associated with sustainable energy practices and environmental stewardship.

g) "Energy Intensity per Capita" involves assessing the amount of energy consumed per person in a given country or region. This metric provides valuable insights into the efficiency of energy use on an individual basis and has several implications

h) Energy intensity by GDP serves as an indicator of how efficiently a country or region produces economic output with its energy resources. Lower energy intensity values suggest greater economic productivity for each unit of energy consumed.

i) Examining historical CO2 emissions data enables the identification of trends over time. This analysis helps understand how emissions have evolved, which is crucial for assessing the effectiveness of environmental policies, technological advancements, and societal changes

5.Data Visualization:

Data visualization is a critical part of Exploratory Data Analysis (EDA), and you can create informative and visually appealing plots using Python libraries like Matplotlib and Seaborn.

Graphs and plots used through Matplotlib and Seaborn

Create Visual Summaries: Generate summary statistics (e.g., mean, median, standard deviation). Create histograms, box plots, and density plots to visualize the data distribution

Explore Relationships: Use scatter plots, pair plots, and correlation matrices to explore relationships between variables.

Analyse Categorical Data: Visualize categorical data using bar charts, count plots, and contingency tables

Time Series Analysis: Plot time series data to detect trends, seasonality, and anomalies.

Bar Plot (Grouping by State):A bar plot is used to visualize categorical data.

Scatter Plot: A scatter plot is used to visualize the relationship between two numerical variables. We can create a scatter plot to compare

Line Plot (Time Series Analysis):A line plot is useful for visualizing time series data. We can use this to create a time series analysis

Pair Plot (Multiple Variables Comparison):A pair plot is used for exploring relationships between multiple variables simultaneously. Here, we can create a pair plot using variables

Box Plot (Outlier Detection):A box plot is useful for identifying outliers and visualizing the distribution of data

Univariate analysis:

Univariate analysis is a statistical method used to analyze the distribution, central tendency, and variability of a single variable at a time. It focuses on understanding the characteristics and patterns of a single variable without considering relationships with other variables. Univariate analysis is often

the first step in statistical analysis and is useful for summarizing and describing the main features of a dataset
Columns used:CO2 emission
Graphs used: distplot,subplot

BIVARIATE ANALYSIS:

Bivariate analysis involves the examination of the relationship between two variables. The goal is to understand the patterns, trends, and associations between these two variables. The goal is to understand how changes in one variable are related to changes in another.Here are some common techniques used in bivariate analysis
Columns used: e_con, e_type,co2 emission,Year
Graphs used: Pie chartline plot ,subplot,stripplot,bar chart,bubbleplot

MULTIVARIATE ANALYSIS:

Multivariate analysis involves the examination of the relationships among three or more variables. It goes beyond the scope of bivariate analysis and allows for a more comprehensive understanding of how multiple variables interact
Columns Used:Year,country,e-con,e-prod,energy intensity per capita,energy intensity by gdp,gdp
Graphs used: Heatmaps

Questions for Analysis:

1. What are the names and data types of the columns?
2. What are the basic summary statistics?
3. Are there any categorical variables and missing values? If so print it?
4. What are the units of measurement for numerical columns? (example: time , currency, date, distance) ?
5. Do you have domain clarification? Brief it?
6. Are there any correlations between variables? Calculate correlations?
7. What is yearly energy consumption for each energy type?
8. ..Is there any yearly percentage change in energy consumption?
9. Show each change in percentage of each energy type?
10. Top energy consumers over the time period?
11. what energy types do the top 4 consumers use?

12 Total Consumption of each energy type for top 4 consumers over entire period.?

13. Who consumes the most clean/dirty energy worldwide?

14 CO2 Emission, Energy Intensity by GDP and Energy Intensity per Capita of top 4 consumers?

15.Relationship between energy production and CO2 emission globally?

16.Top energy producers over the time period?

17.What energy types do top energy producers make?

18. Total Produced of each energy type for top 4 producers over entire period?

19. Who produces the most clean/dirty energy worldwide?

20. CO2 Emission, Energy Intensity by GDP, and Energy Intensity per Capita of top 4 producers?

21.what is the relationship of globally co2 emission and energy consumption?

22.Are there any percentage change in co2 emission each year?

23. Yearly Mean, Median and Standard Deviation of each countries CO2 Emission?

24. Who produced the most CO2 Emissions over this time period?

25. Yearly CO2 emissions of top 6 Countries over the time period?

26. Countries with the most drastic increase/decrease in CO2 emission from the start to end of this time period?

27. Top Countries with the largest Population and there CO2 emissions as of 2019?

28. Top Countries with the highest GDP and lowest CO2 emissions as of 2019 (Best economy with lowest pollution)?

29. Top Countries with the lowest Energy Intensity of GDP and there CO2 emissions as of 2019 (Best energy conservation with lowest pollution)?

30. Top Countries with the largest Energy Intensity per Capita and there CO2 emissions (Best energy conservation per capita with lowest pollution)?

31 Top Countries with the largest Energy Consumption and their CO2 emissions?

Visualization and insights of Questions:

What are the names and data types of the columns?

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 55440 entries, 0 to 55439
Data columns (total 10 columns):
 #   Column         Non-Null Count  Dtype  
---  ------         --------------  -----  
 0   Country        55440 non-null  object 
 1   e_type         55440 non-null  object 
 2   Year           55440 non-null  int64  
 3   e_con          44287 non-null  float64
 4   e_prod         44289 non-null  float64
 5   GDP            40026 non-null  float64
 6   Population     46014 non-null  float64
 7   ei_capita      50358 non-null  float64
 8   ei_gdp         50358 non-null  float64
 9   CO2_emission   51614 non-null  float64
dtypes: float64(7), int64(1), object(2)
memory usage: 4.2+ MB
```

What are the basic summary statistics?

Out[121]:

	Year	e_con	e_prod	GDP	Population	ei_capita	ei_gdp	CO2_emission
count	35863.000000	35863.000000	35863.000000	35863.000000	3.586300e+04	35863.000000	35863.000000	35863.000000
mean	2003.933134	1.591626	1.585807	837.823183	6.814527e+04	82.817229	4.621024	95.511858
std	9.105176	16.270606	16.072447	6212.660437	4.868524e+05	119.402517	4.367149	1027.760136
min	1988.000000	-0.040000	0.000000	0.120000	1.616000e+01	0.000000	0.000000	-0.010000
25%	1996.000000	0.000000	0.000000	10.300000	1.822240e+03	9.450000	2.290000	0.000000
50%	2004.000000	0.020000	0.000000	44.940000	7.010160e+03	40.330000	3.660000	0.060000
75%	2012.000000	0.200000	0.100000	247.210000	2.234660e+04	107.830000	5.590000	7.520000
max	2019.000000	601.040000	611.510000	127690.250000	7.714631e+06	1139.320000	166.910000	35584.930000

Are there any categorical variables and missing values? If so print it?

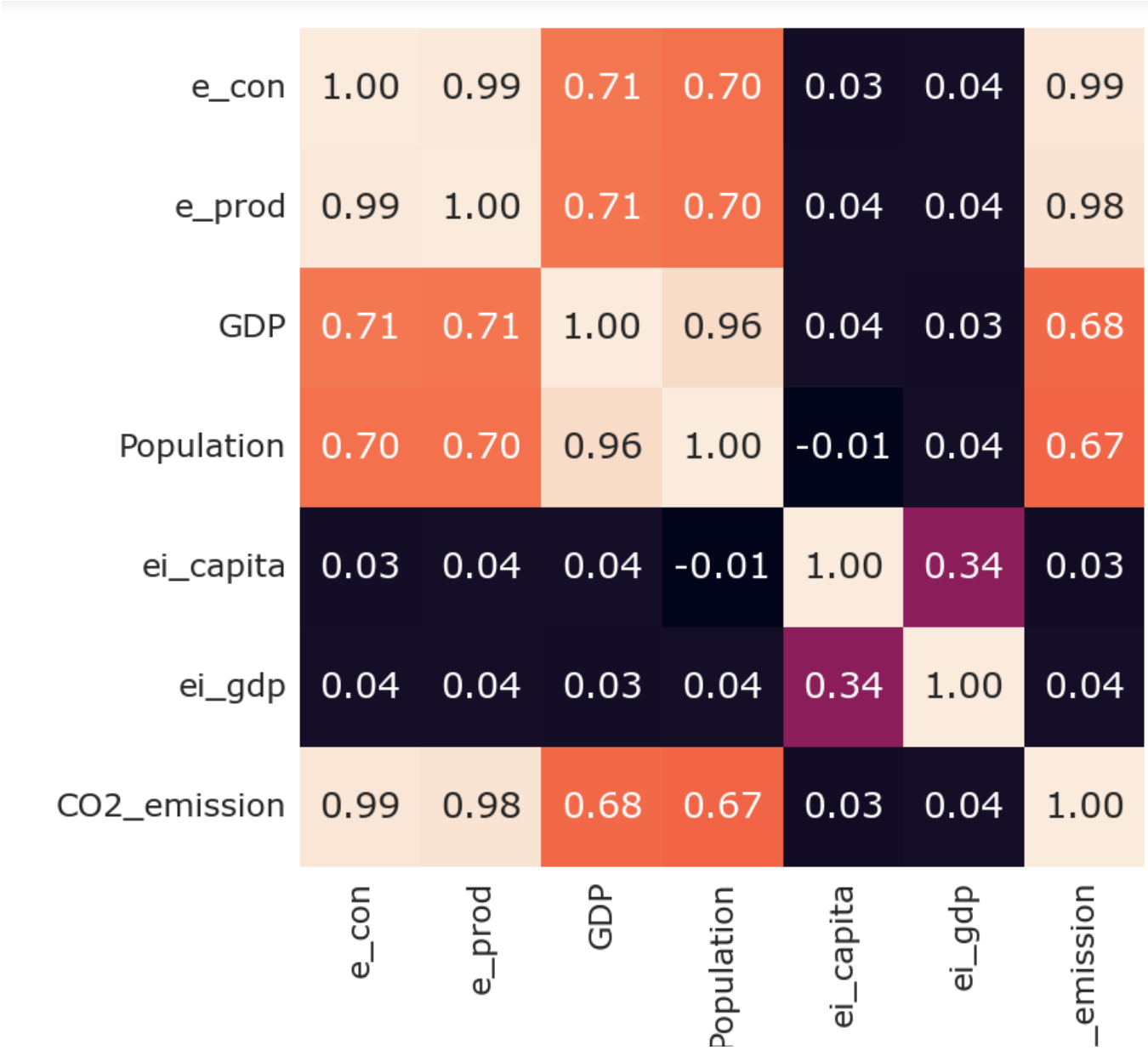
```
11088        all
11089       coal
11090    nat_gas
11091    pet/oth
11092    nuclear
          ...
55435       coal
55436    nat_gas
55437    pet/oth
55438    nuclear
55439    ren/oth
Name: e_type, Length: 35863, dtype: object
```

Do you have domain clarification? Brief it

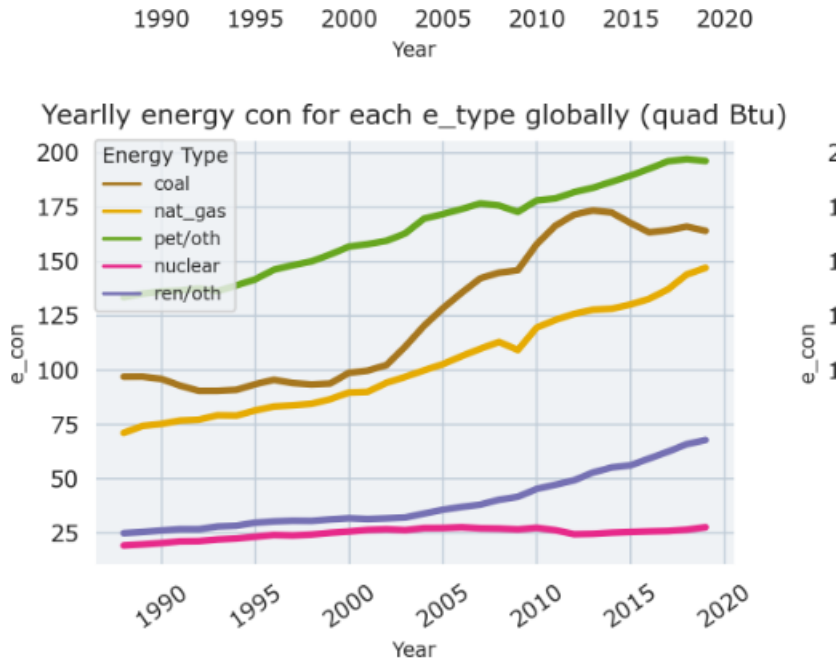
Ans: Exploratory Data Analysis (EDA) on the dataset focusing on "Energy Consumption/Production and CO2 Emission around the World" involves a systematic examination to derive valuable insights into the complex interrelationships between global energy dynamics and environmental impact. The initial steps include a thorough understanding of the dataset's structure, encompassing variables such as energy consumption, production, CO2 emissions, GDP, and population figures. Descriptive statistics offer an initial overview of the data, paving the way for a detailed temporal analysis that identifies trends over time. Geographic exploration reveals regional disparities and global patterns, shedding light on countries or regions with significant contributions to energy consumption and CO2 emissions. Sectoral breakdowns allow for a nuanced examination of energy usage, while per capita metrics provide insights into individual contributions. The analysis extends to the economic realm, assessing the correlation between GDP and energy metrics. Visualization techniques, including time series plots and geospatial representations, facilitate a more intuitive understanding of the dataset. This comprehensive EDA not only aids in recognizing historical trends but also serves as a foundation for generating hypotheses, identifying policy implications, and contributing to a holistic understanding of the intricate relationship between global energy dynamics and environmental sustainability.

Are there any correlations between variables? Calculate correlations?

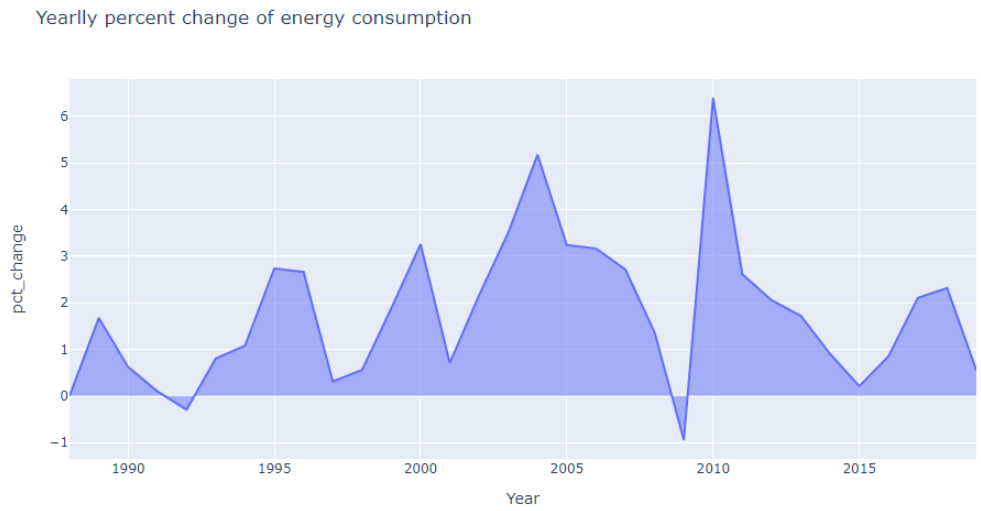
	Year	e_con	e_prod	GDP	Population	ei_capita	ei_gdp	CO2_emission
Year	1.000000	0.016574	0.016557	0.037829	0.011503	0.040577	-0.090976	0.015474
e_con	0.016574	1.000000	0.993353	0.712189	0.698024	0.034380	0.039014	0.985196
e_prod	0.016557	0.993353	1.000000	0.708141	0.697747	0.040172	0.043559	0.977733
GDP	0.037829	0.712189	0.708141	1.000000	0.957655	0.037593	0.033350	0.679897
Population	0.011503	0.698024	0.697747	0.957655	1.000000	-0.014916	0.038245	0.670486
ei_capita	0.040577	0.034380	0.040172	0.037593	-0.014916	1.000000	0.339646	0.029050
ei_gdp	-0.090976	0.039014	0.043559	0.033350	0.038245	0.339646	1.000000	0.038943
CO2_emission	0.015474	0.985196	0.977733	0.679897	0.670486	0.029050	0.038943	1.000000



What is yearly energy consumption for each energy type?

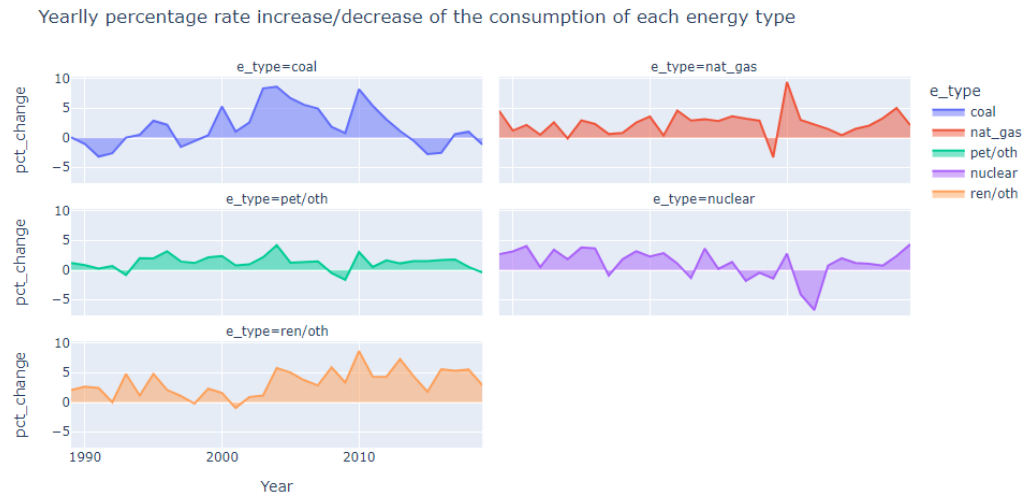


Is there any yearly percentage change in energy consumption?

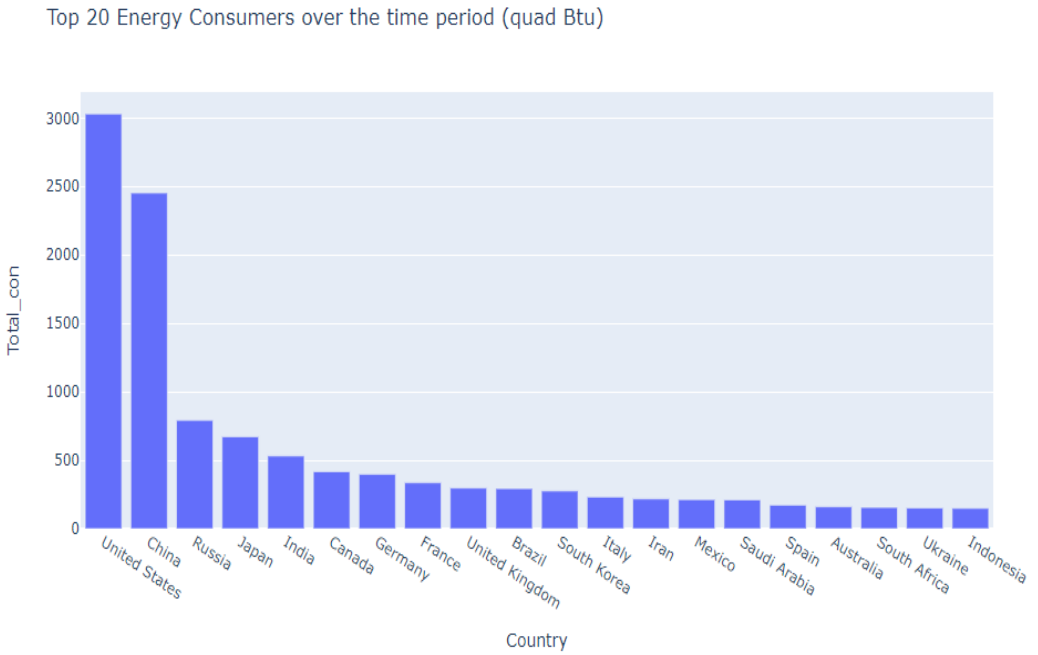



```
The average yearly percentage change is 1.81%,
 and the overall percentage change from 1988 to 2019 is 73.93%
```

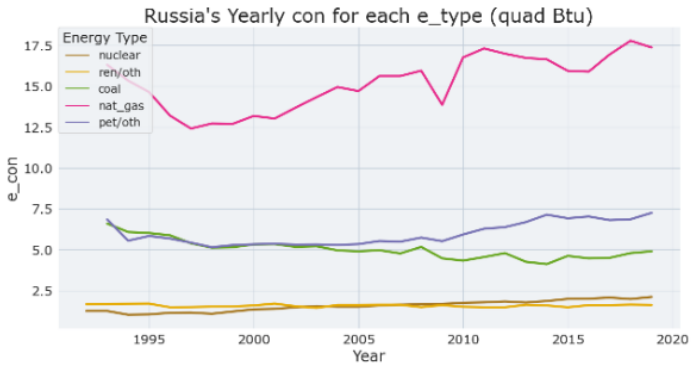
Show each change in percentage of each energy type?

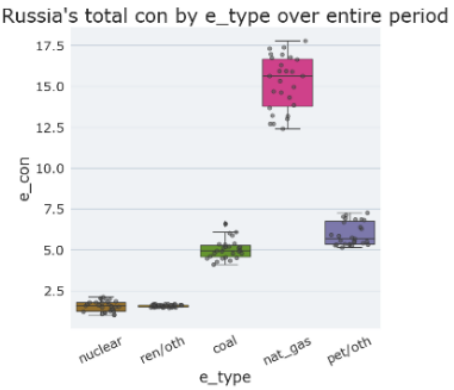


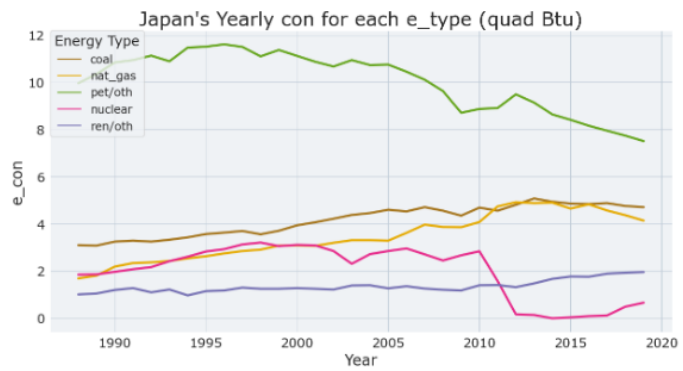
Top energy consumers over the time period?

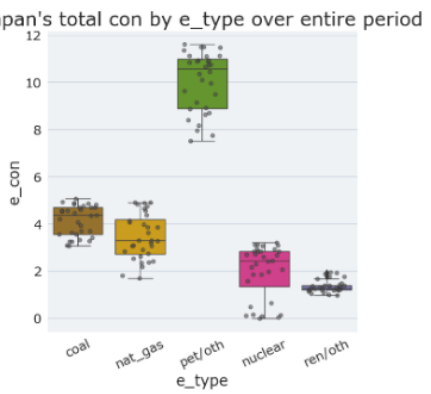


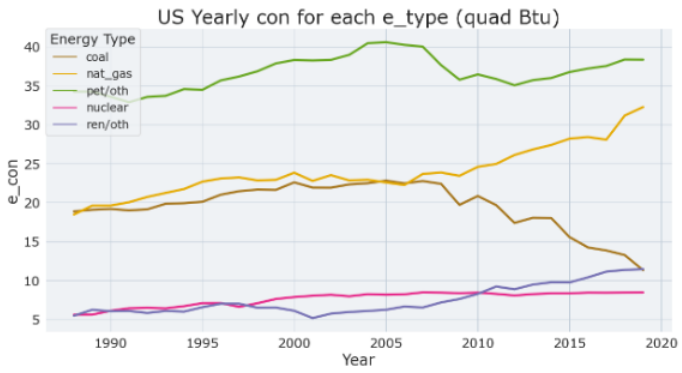
what energy types do the top 4 consumers use

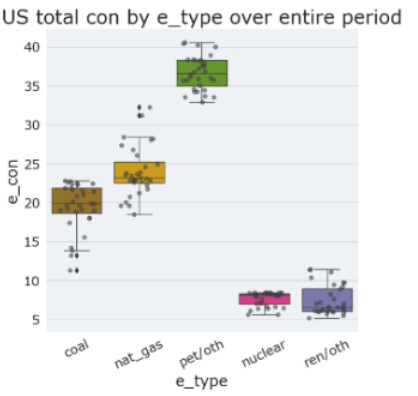


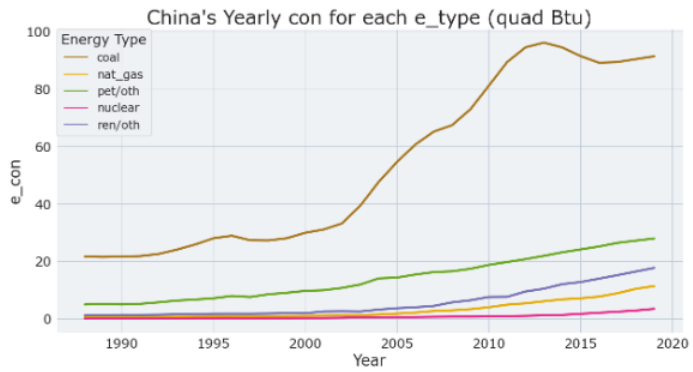


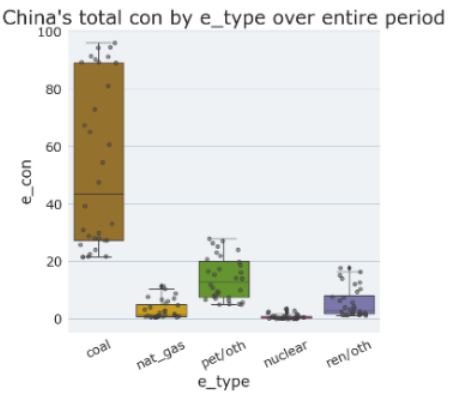




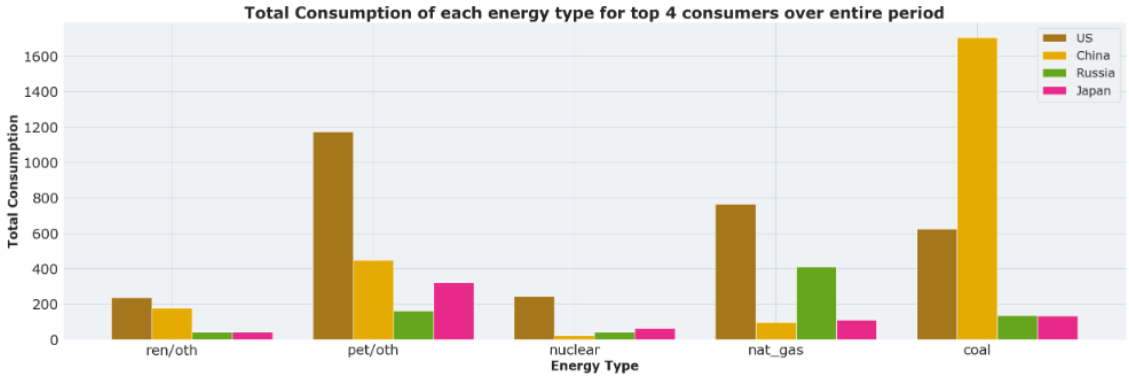




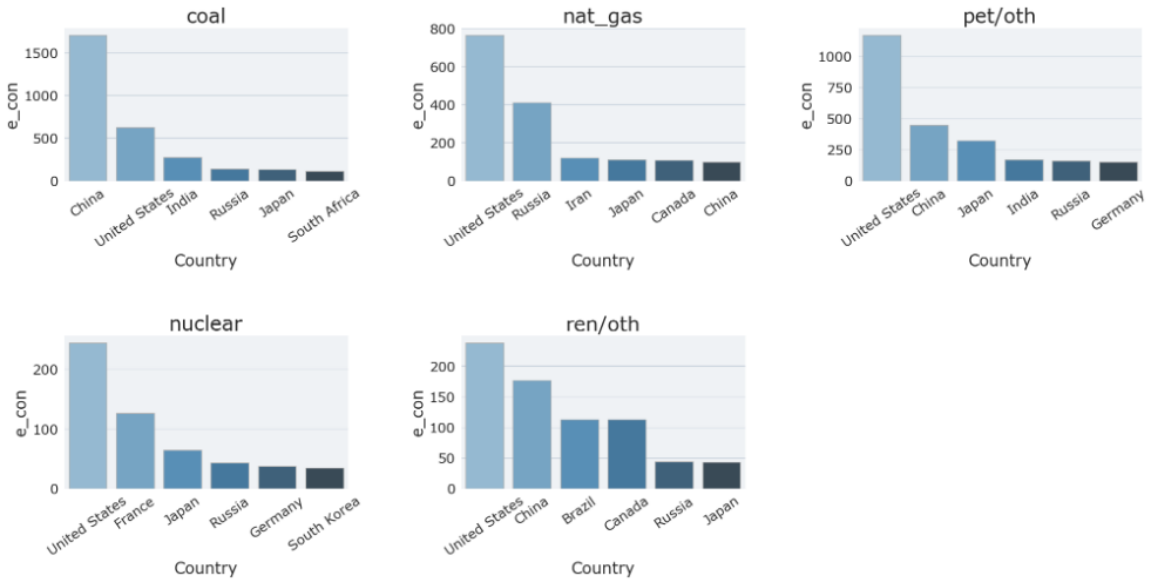


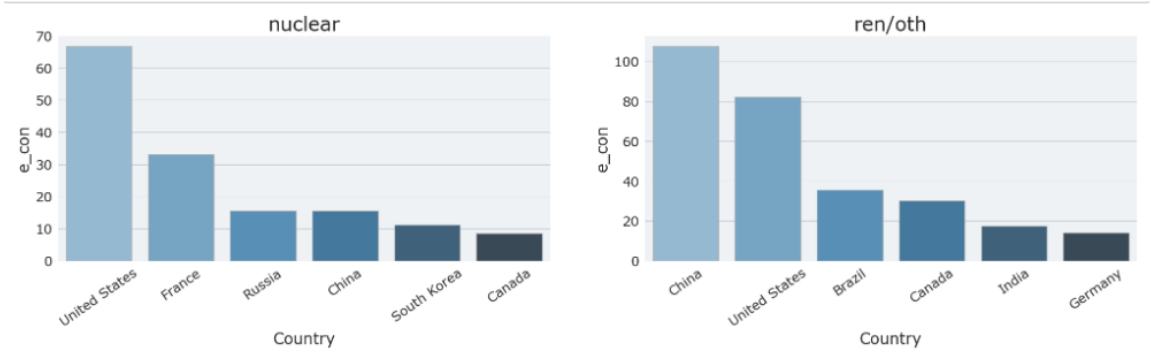


Total Consumption of each energy type for top 4 consumers over entire period.?



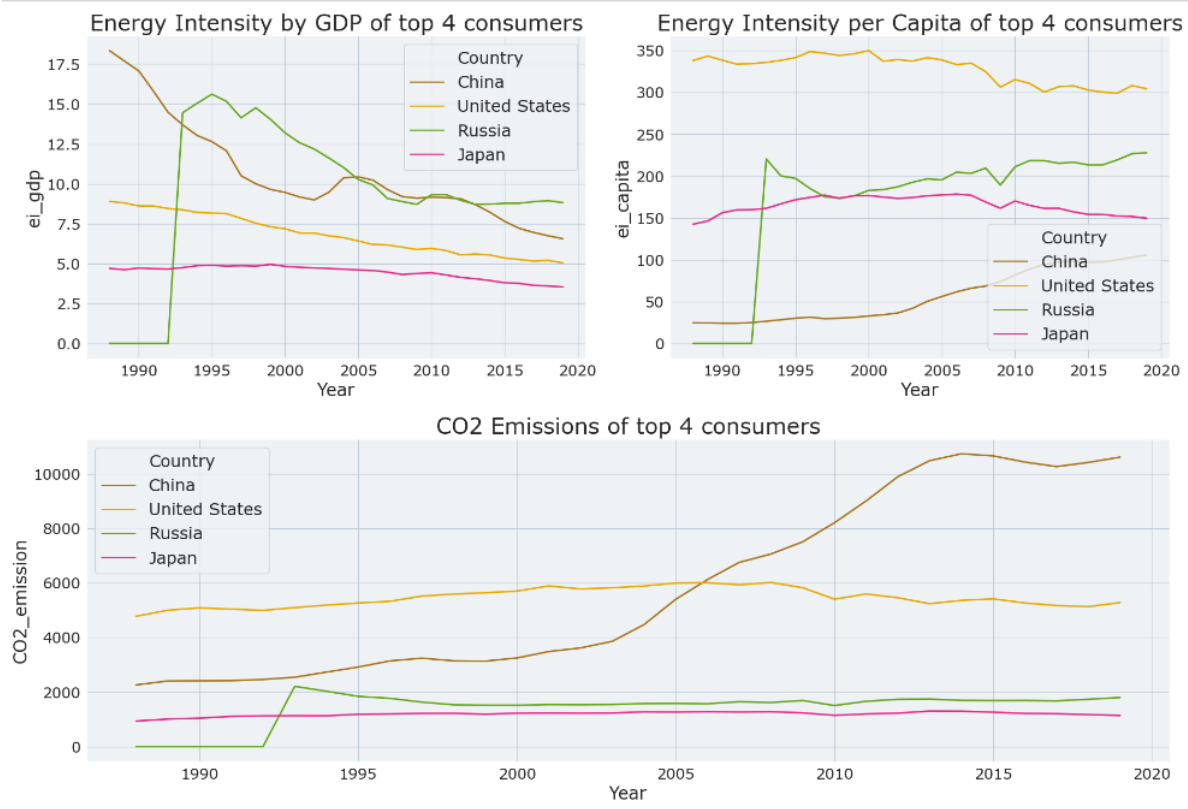
. Who consumes the most clean/dirty energy worldwide?



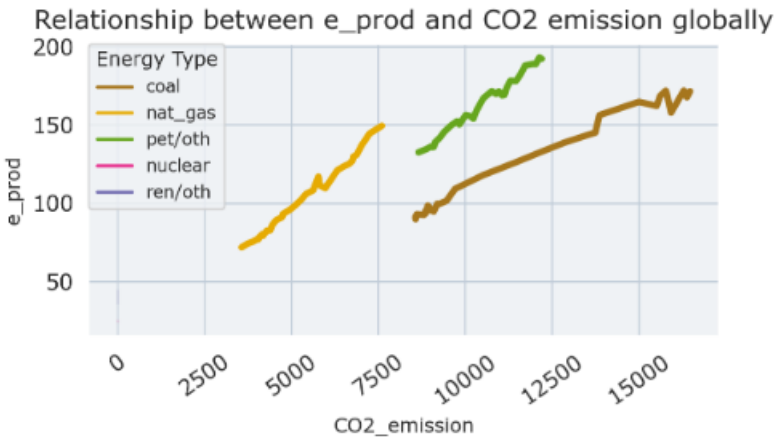


The United States is still the dominant consumer of nuclear energy, but as for renewables/other, China seems to have increased its consumption and surpassed The United States in becoming the dominant consumer of renewables.

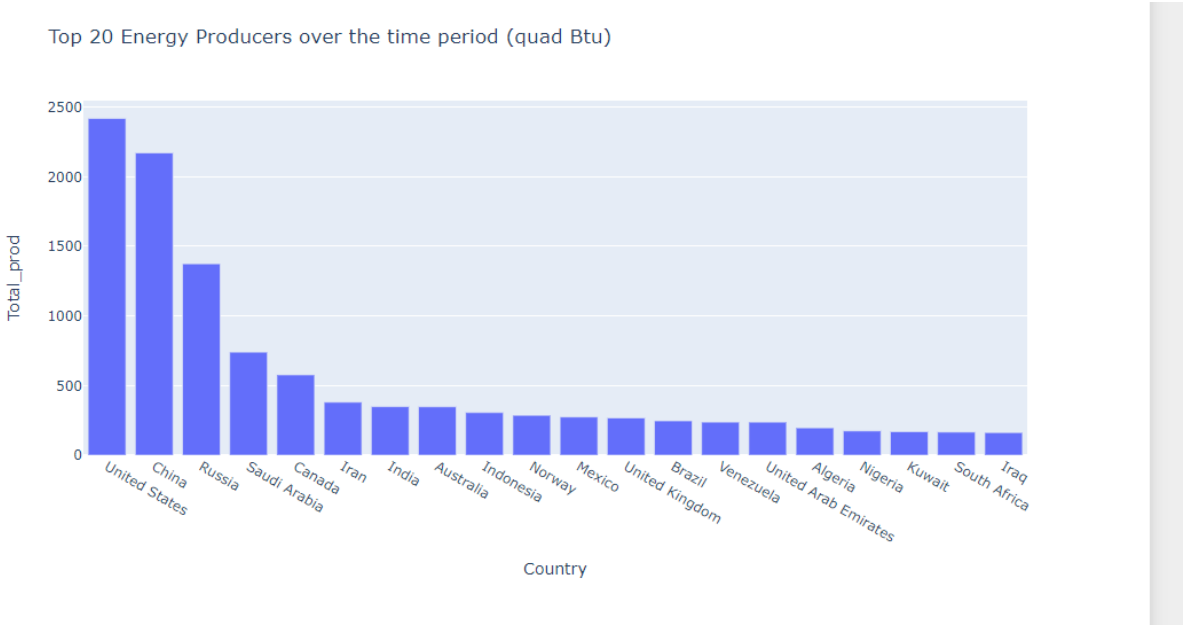
CO2 Emission, Energy Intensity by GDP and Energy Intensity per Capita of top 4 consumers?



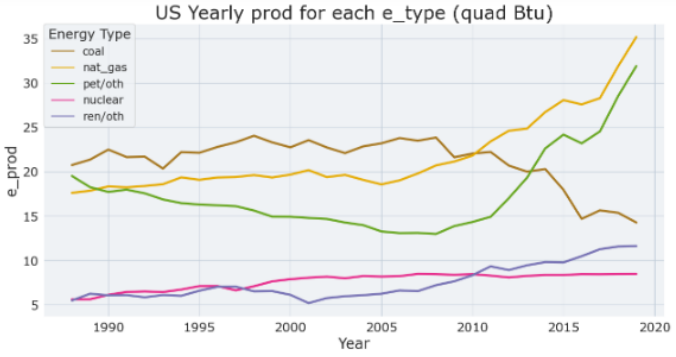
.Relationship between energy production and CO2 emission globally?

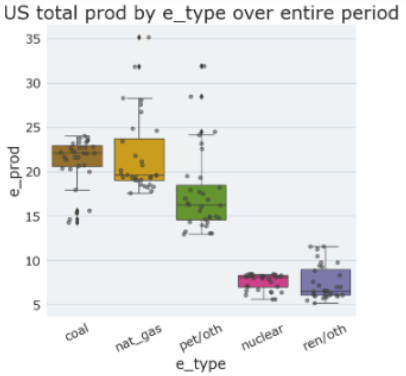


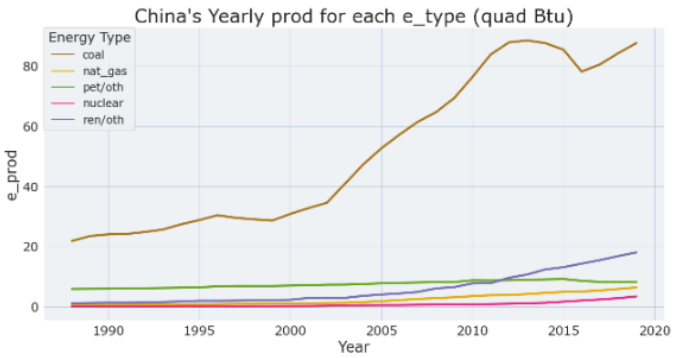
Top energy producers over the time period?

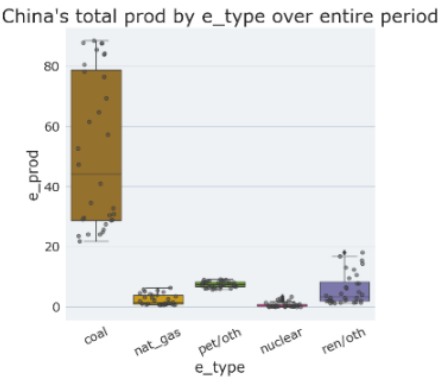


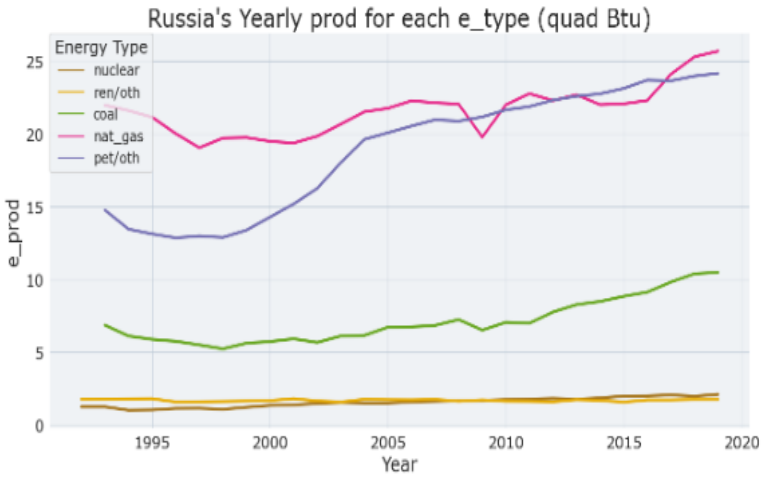
What energy types do top energy producers make?

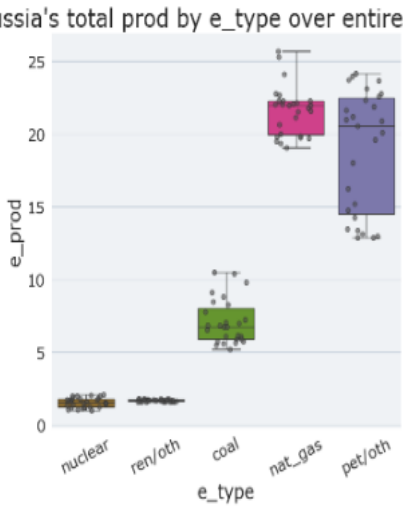


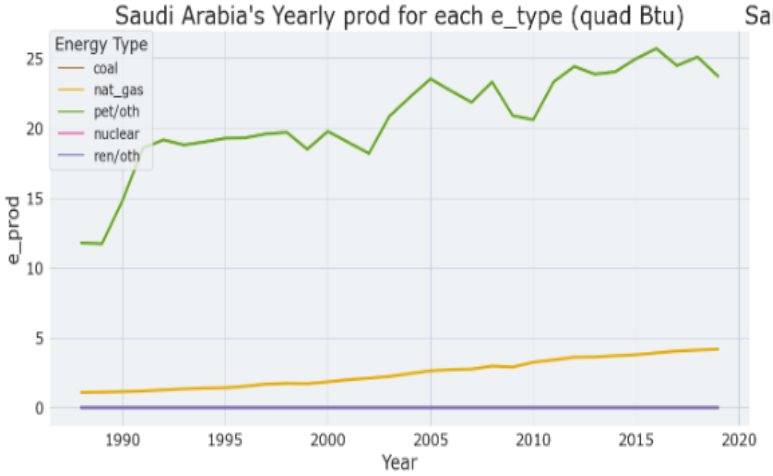


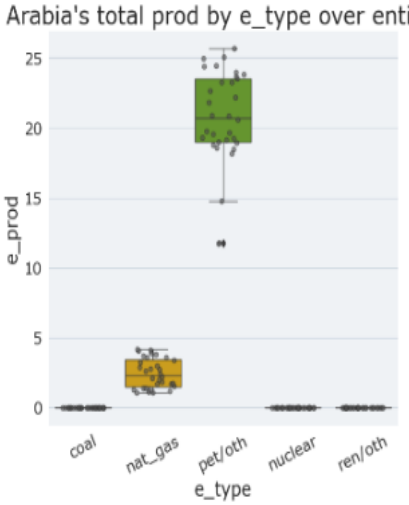




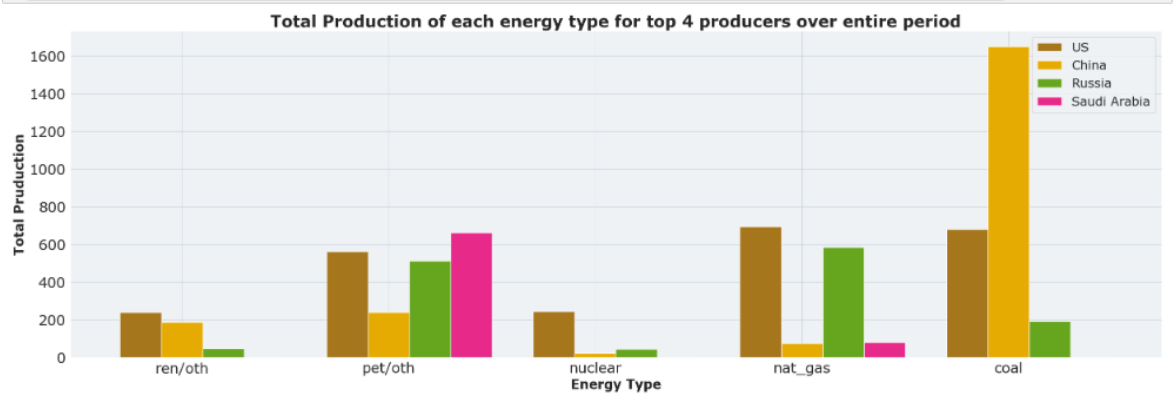




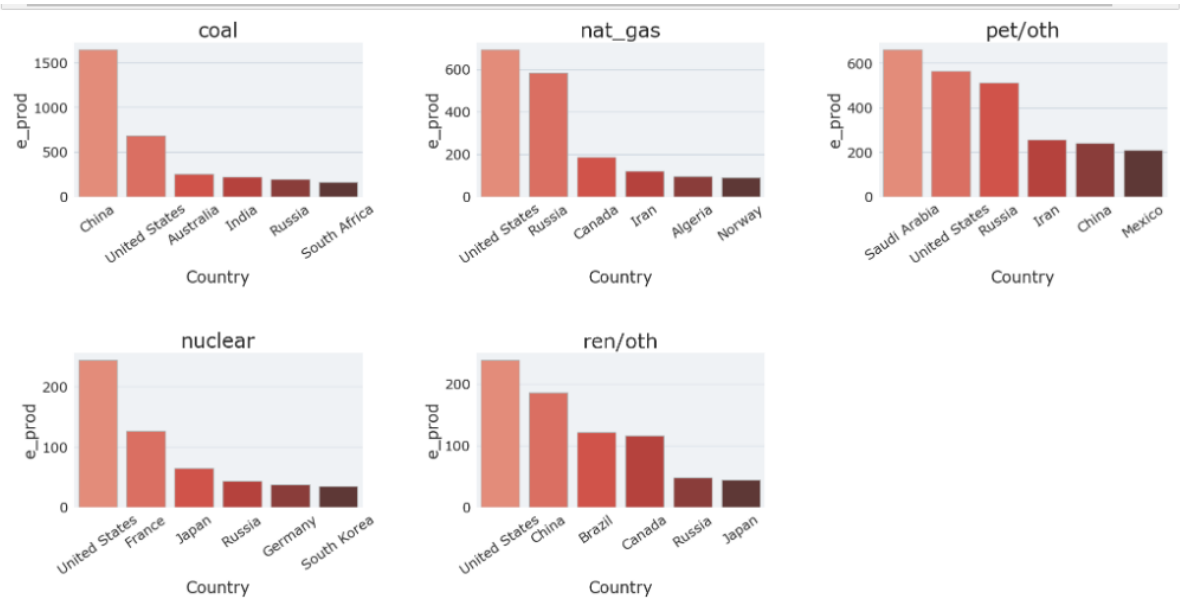




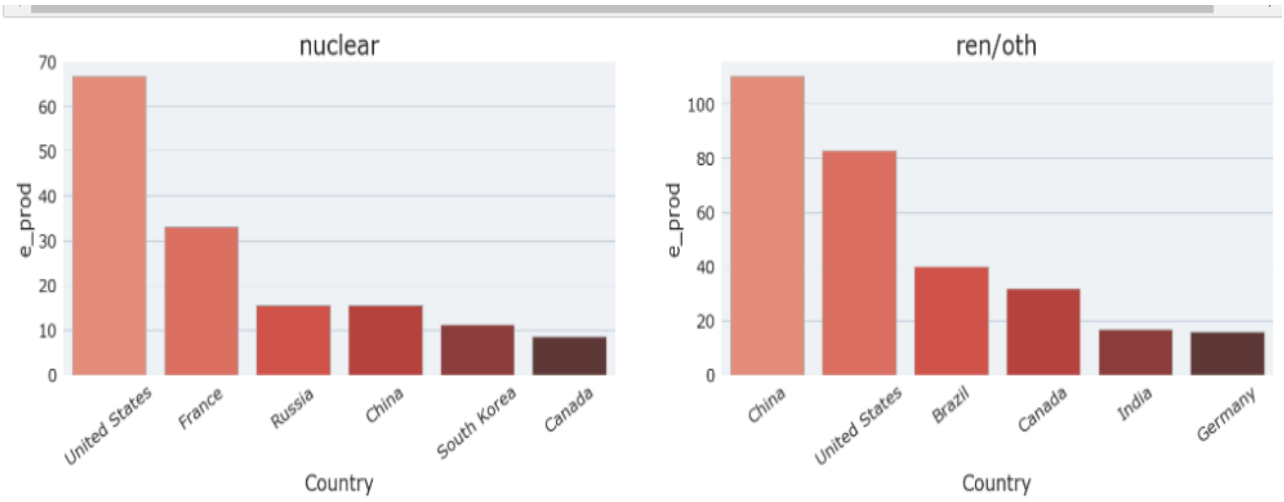
. Total Produced of each energy type for top 4 producers over entire period?



. Who produces the most clean/dirty energy worldwide?

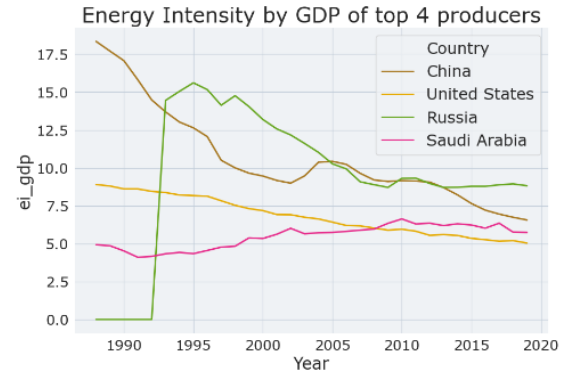


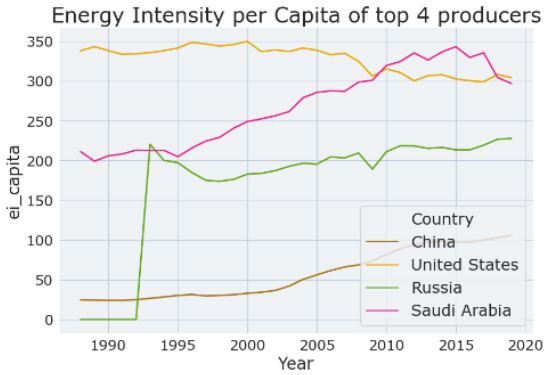
Seems that over the entire time period, The United States has been the dominant producer of clean energy.

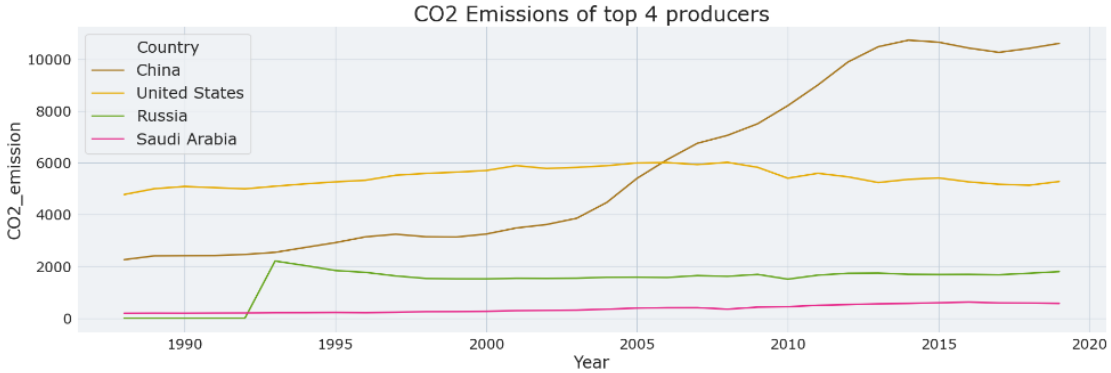


The United States is still the dominant producer of nuclear energy, but as for renewables/other, China seems to have picked up its production and surpassed The United States in becoming the dominant producer of renewables.

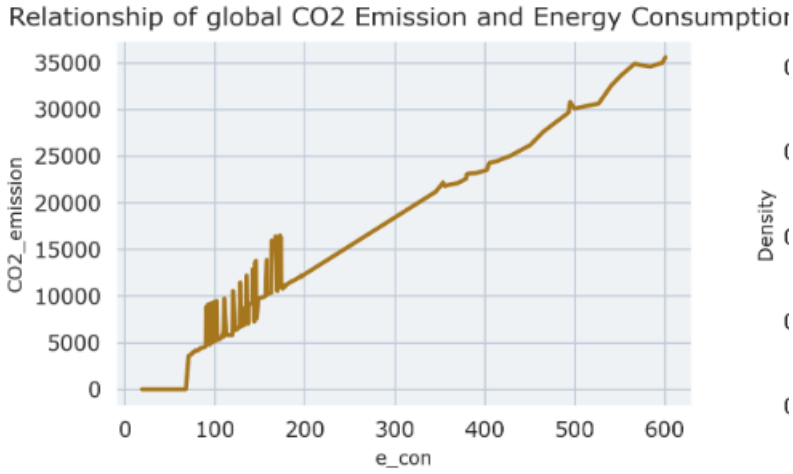
CO2 Emission, Energy Intensity by GDP, and Energy Intensity per Capita of top 4 producers?



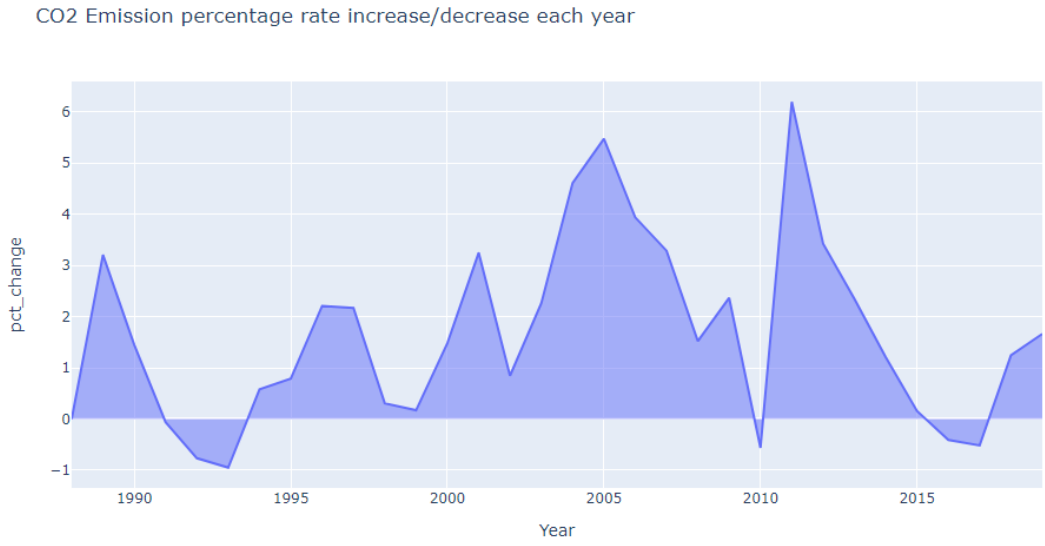




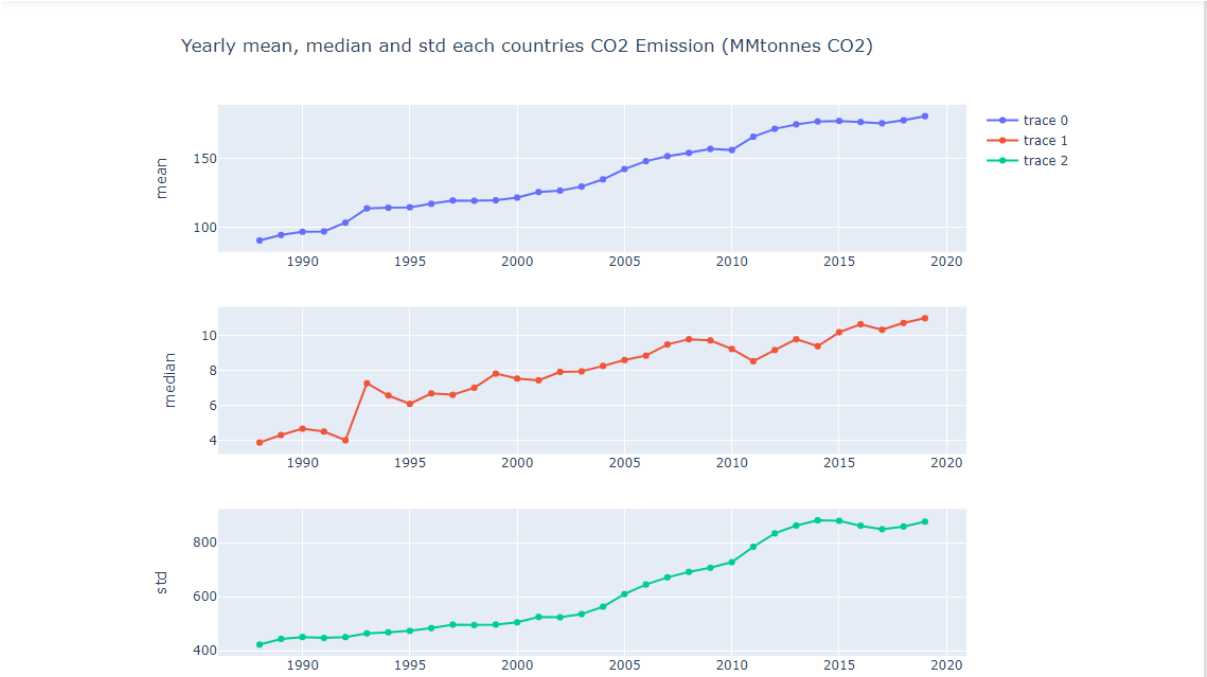
what is the relationship of globally co2 emission and energy consumption?



Are there any percentage change in co2 emission each year?

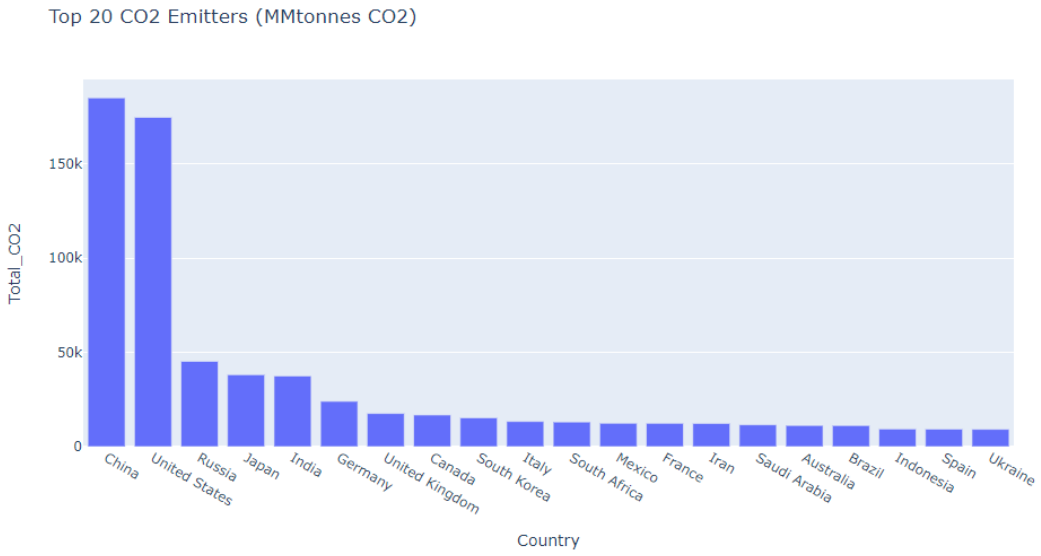


. Yearly Mean, Median and Standard Deviation of each countries CO2 Emission?



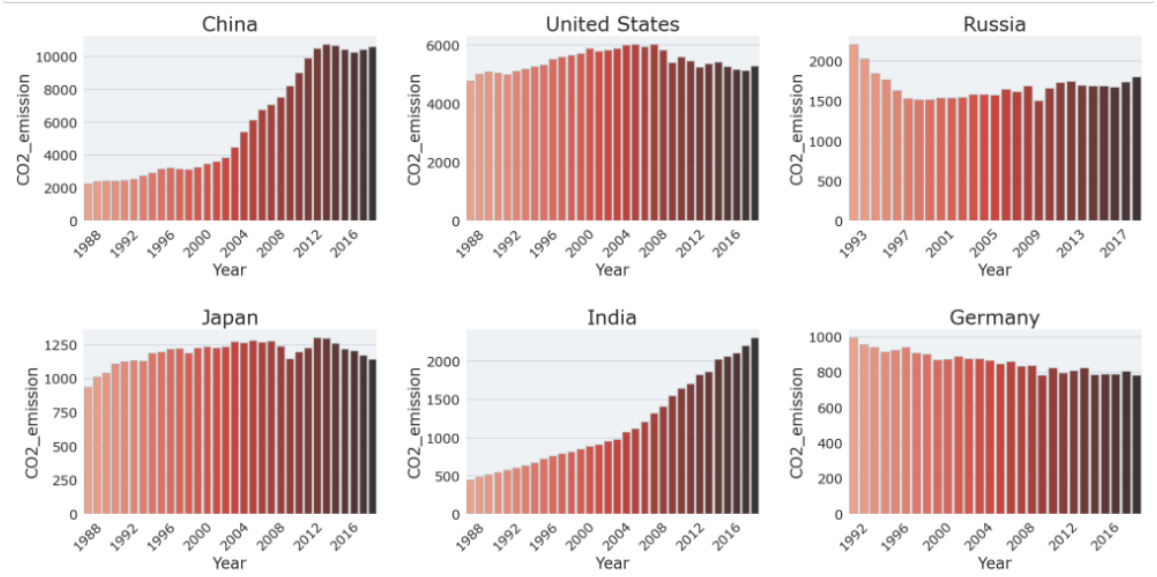

```
The worlds Mean value of CO2 Emission over the entire time period is 27434.91.
The worlds Median value of CO2 Emission over the entire time period is 25592.93.
The Standard Deviation value of CO2 Emission over the entire time period is 5256.78.
```

Who produced the most CO2 Emissions over this time period?



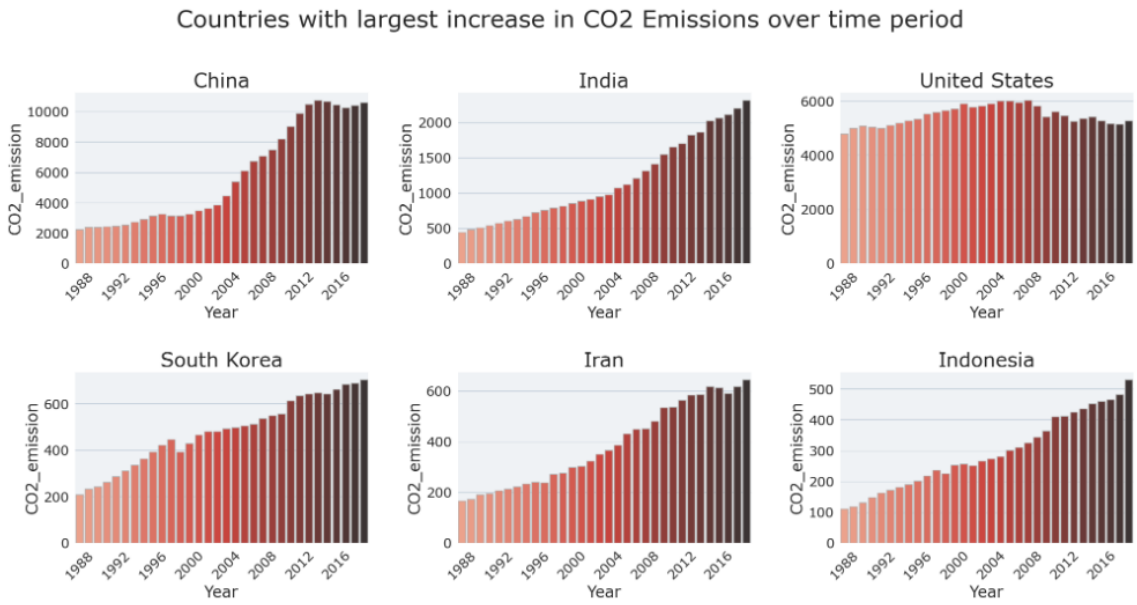
The top CO2 emitters are China and The US, both exceding nearly 4 times or more the amount of every other country!

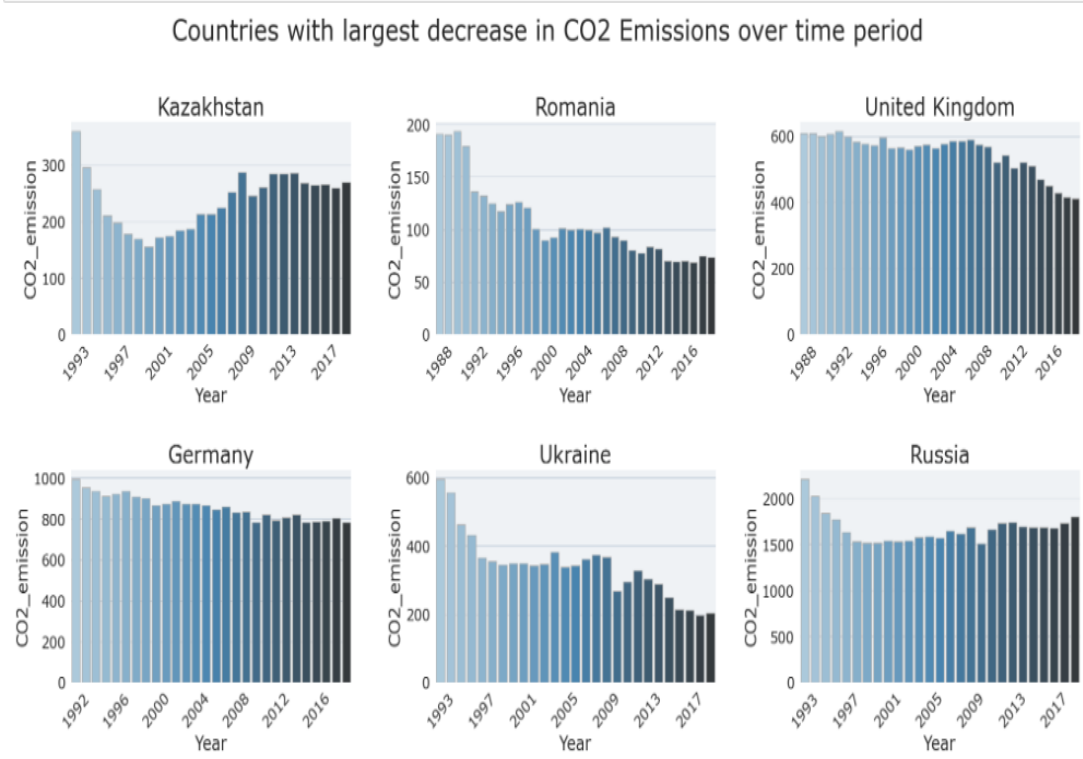
Yearly CO2 emissions of top 6 Countries over the time period?



China and India have increased there emissions a lot over the time period.

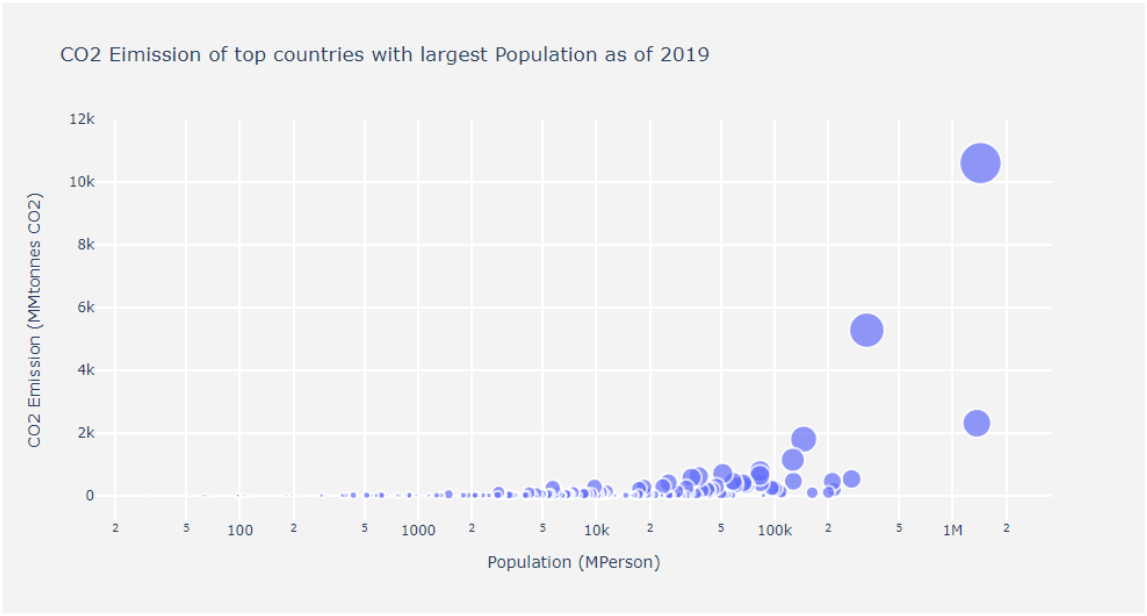
Countries with the most drastic increase/decrease in CO2 emission from the start to end of this time period?





The countries with the largest decrease in CO2 Emission are all mostly former soviet republics, I assume after the USSR fell there production/industry fell drastically, which in return decreased there emissions. Germany (technically apart of USSR, techinically not) and The United Kingdom have also decreased there emissions quite a bit over the time period.

Top Countries with the largest Population and there CO2 emissions as of 2019?



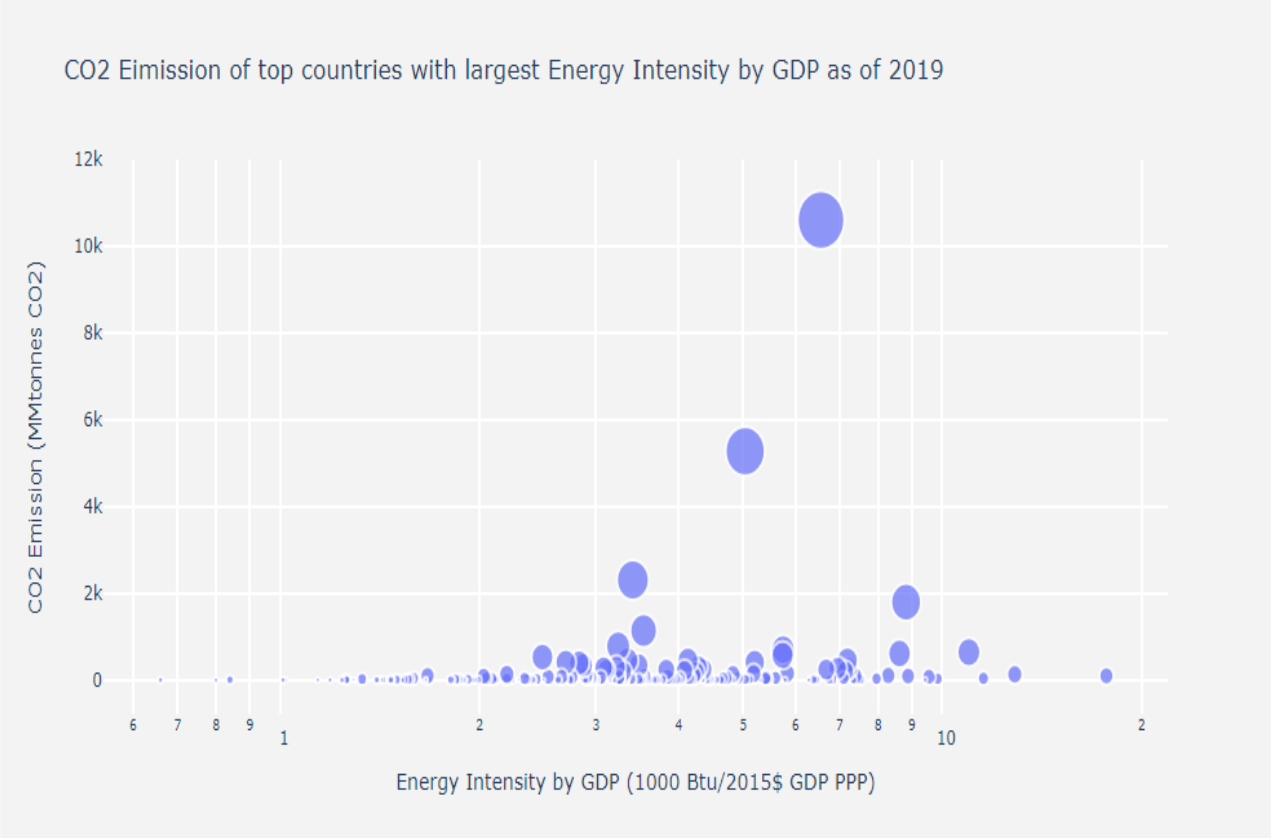
Generally speaking, the larger the population, the more CO2 the country will be likely to emit

Top Countries with the highest GDP and lowest CO2 emissions as of 2019 (Best economy with lowest pollution)?

	Country	Year	GDP	CO2_emission
0	China	2019	23128.34	10608.60
1	United States	2019	19925.43	5277.19
2	India	2019	9310.29	2308.33
3	Japan	2019	5359.05	1140.95
4	Germany	2019	4182.05	781.60
5	Russia	2019	3770.42	1798.24
6	Indonesia	2019	3227.67	529.40
7	Brazil	2019	3041.30	453.44
8	United Kingdom	2019	2992.82	411.17
9	France	2019	2918.62	341.98

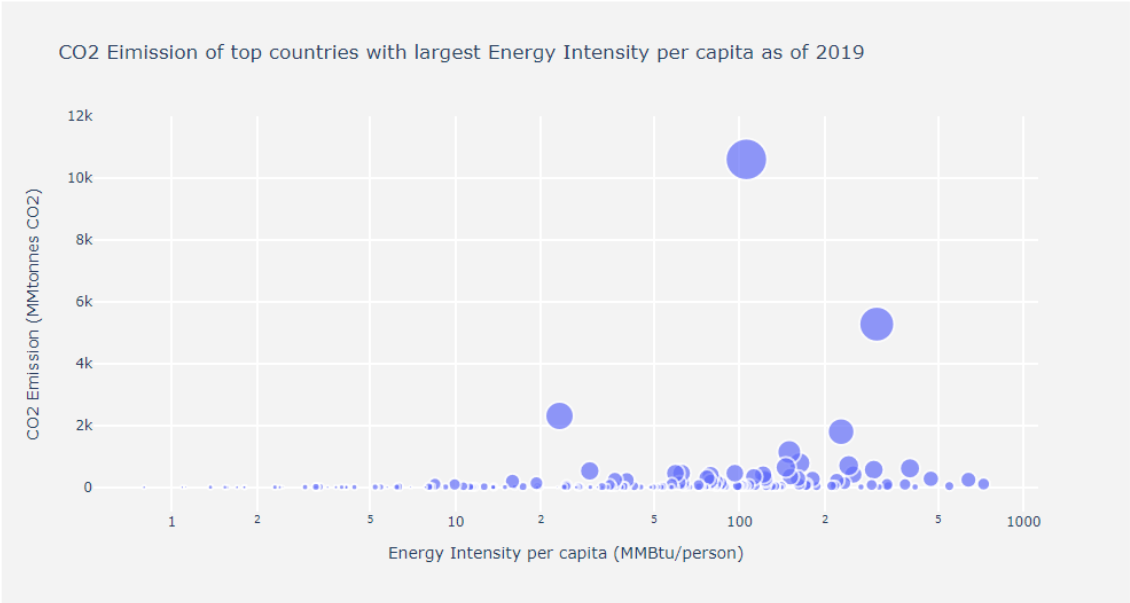
Top Countries with the lowest Energy Intensity of GDP and there CO2 emissions as of 2019 (Best energy conservation with lowest pollution)?

	Country	Year	ei_gdp	CO2_emission
0	Turkmenistan	2019	17.70	97.31
1	Venezuela	2019	12.87	131.49
2	Bahrain	2019	11.54	39.01
3	Iran	2019	10.97	642.96
4	Laos	2019	9.84	29.95
...	...	...	...	...
188	Sierra Leone	2019	1.02	0.99
189	Chad	2019	1.01	1.72
190	Zambia	2019	0.84	7.10
191	Rwanda	2019	0.80	1.11
192	Macau	2019	0.66	2.01



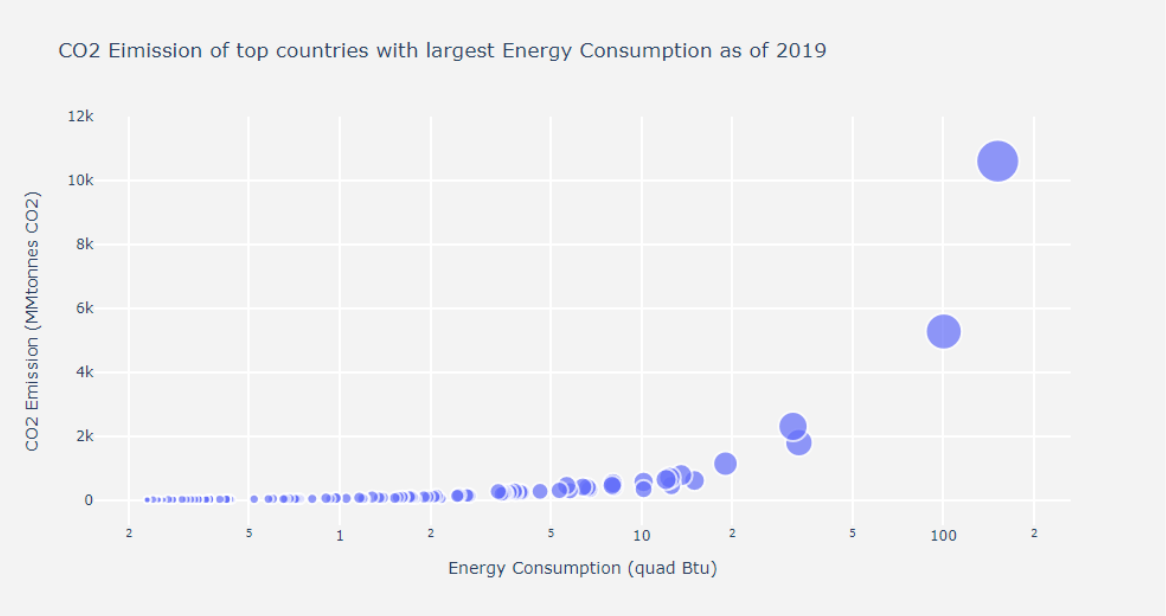
A low/high Energy Intensity by GDP isnt necesarily predictive of a large CO2 Emission.

Top Countries with the largest Energy Intensity per Capita and there CO2 emissions (Best energy conservation per capita with lowest pollution)?



A low/high Energy Intensity per capita isnt necesarily predictive of a large CO2 Emission.

Top Countries with the largest Energy Consumption and their CO2 emissions?



Findings and insights:

The findings and insights from an Exploratory Data Analysis (EDA) on a dataset related to "Energy Consumption/Production, CO2 Emission around the World" can be diverse and multifaceted. Here are potential findings and insights that might emerge from such an analysis:

1. **Temporal Trends:**

- Identification of temporal trends in energy consumption, production, and CO2 emissions over the years.
- Recognition of periods of significant growth or decline in energy-related metrics.

2. **Geographic Patterns:**

- Identification of countries or regions with the highest and lowest levels of energy consumption, production, and CO2 emissions.
- Exploration of regional disparities and variations in the energy mix.

3. **Energy Mix Composition:**

- Understanding the composition of the global energy mix, including the share of fossil fuels, renewables, and nuclear energy.
- Recognition of shifts in energy production patterns and the growing role of renewable energy sources.

4. **CO2 Emission Hotspots:**

- Identification of major sources and sectors contributing to CO2 emissions.
- Assessment of the correlation between energy production methods and CO2 emissions.

5. **Per Capita Analysis:**

- Insights into how individual energy consumption and CO2 emissions vary on a per capita basis.
- Recognition of countries with high per capita energy usage or emissions.

6. **Economic Correlations:**
   - Understanding the relationship between GDP and energy metrics, assessing the energy intensity of economic activities.
   - Identification of economically efficient or inefficient energy users.
7. **Sustainability Assessment:**
   - Evaluation of trends in energy intensity, indicating improvements or declines in energy efficiency.
   - Recognition of countries progressing towards decoupling economic growth from CO2 emissions.
8. **Policy Implications:**
   - Identification of countries where energy-related policies have had a notable impact on reducing emissions or improving efficiency.
   - Insights into potential areas for policy interventions to achieve environmental sustainability goals.
9. **Renewable Energy Adoption:**
   - Recognition of countries or regions leading in the adoption of renewable energy sources.
   - Assessment of the correlation between renewable energy use and reduced CO2 emissions.
10. **Global Comparisons:**
    - Comparative analysis of energy metrics and emissions across different continents or economic groupings.
    - Insights into international benchmarks and opportunities for collaborative efforts.
11. **Future Projections:**
    - Exploration of any available projections or forecasts regarding future energy consumption, production, and CO2 emissions.
    - Insights into potential trajectories and challenges for achieving global sustainability goals.

12. **Sectoral Breakdown:**

- Identification of sectors contributing significantly to energy consumption and emissions.
- Recognition of industries where targeted interventions may be most effective.

Limitations:

1. Understanding complex real-world data sets
2. Handling missing values in numerical and categorical data
3. Difficult to judge whether to remove or replace missing values and outliers

4.Converting data types of columns

5.Difficulties in choosing graphs between many, for the exact representation of data

6.Difficulty in remembering tools for data visualization

7.Some difficulties while writing Python programming for extraction of required data

Recommendations:

1. **Diversification of Energy Sources:**
   - Encourage a shift towards a diversified and sustainable energy mix that incorporates a higher share of renewable energy sources, reducing reliance on fossil fuels.
2. **Energy Efficiency Initiatives:**
   - Implement and promote energy efficiency initiatives in industries, transportation, and households to reduce overall energy consumption and minimize waste.
3. **Policy and Regulatory Frameworks:**
   - Establish and enforce robust policy and regulatory frameworks that incentivize clean energy production, penalize excessive emissions, and encourage sustainable practices.
4. **Investment in Research and Development:**
   - Allocate resources for research and development in clean energy technologies to enhance efficiency, affordability, and accessibility of renewable energy sources.
5. **International Collaboration:**

- Foster international collaboration on environmental issues, sharing best practices, technologies, and expertise to address global challenges collectively.

6. **Carbon Capture and Storage (CCS):**
   - Invest in and promote the development of Carbon Capture and Storage technologies to mitigate emissions from industries with significant carbon footprints.
7. **Public Awareness and Education:**
   - Raise public awareness about the environmental impact of energy consumption and production.
   - Promote energy literacy to encourage informed choices by individuals and businesses.
8. **Incentives for Sustainable Practices:**
   - Provide financial incentives, tax credits, or subsidies for businesses and individuals adopting sustainable energy practices and reducing their carbon footprint.
9. **Transition to Electric Vehicles:**
   - Support the transition to electric vehicles and invest in the development of efficient and sustainable public transportation systems.
10. **Carbon Pricing:**
    - Consider implementing carbon pricing mechanisms to internalize the environmental costs of carbon emissions, encouraging businesses to reduce their carbon footprint.
11. **Smart Cities and Infrastructure:**
    - Invest in smart city initiatives and infrastructure that promote energy-efficient technologies, renewable energy integration, and sustainable urban development.
12. **Green Building Standards:**
    - Enforce and promote green building standards to reduce energy consumption in the construction and operation of buildings.
13. **Community Engagement:**
    - Encourage community engagement in sustainable practices through initiatives such as community gardens, local renewable energy projects, and waste reduction programs.
14. **Transition Plans for High-Emission Industries:**
    - Develop transition plans for industries with high emissions, facilitating their shift towards cleaner production methods and technologies.
15. **Monitoring and Reporting:**
    - Establish transparent and standardized monitoring and reporting mechanisms for energy consumption, production, and emissions to track progress and identify areas for improvement.

Conclusions:

- Energy prodcution of fossil fuels has a positive relationshiop to CO2 emission, obviouslly...
- Petroleum/other seems to be the dominant energy source, and then coal, natural gas, renewables/other and nuclear following respectively
- World energy production has been increasing around 1.85% yearlly, and has increased 76.02% over the entire time period
- Some of the top producers include The United States, China, Russia and Saudi Arabia
- Most of the top producers have either a large economy or high standard of living
- Over the entire time period, The United States has produced the most Nuclear and Renewable/other energy
- In the past 10 years, The United States is still the dominant producer of Nuclear energy, but China has surpassed the United States and has now become the dominant producer of renewables/other energy
- Energy consumption has a positive relationshiop with CO2 emission
- Most consumed energy types go in the order of Petroleum/other, Coal, Natural Gas, Renwables/other, and lastly Nuclear
- Energy consumption has been increasing by roughly 1.81% yearlly, and has overall increased by 73.93% in this time period
- Top Consumers are The United States, China, Russia, Japan, and India to name a few
- Over the entire time period, The United States has consumed the most clean energy
- In the past ten years The United States is still the dominant consumer of Nuclear, but China has overtaken The United States place in becoming the dominant consumer of Renewable/other energy
- In [ ]:
-
- CO2 Emission has been increasing throughout the time period.
- Coal and Petroleum/other liquids have been the dominant energy source for this time period.
- CO2 Emission has been icreasing 1.71% yearlly on average, and has overall increased by 68.14% over the entire time period.
- As of 2019, the average CO2 emission emitted was 10.98 (MMtonnes CO2) for the year.
- The top CO2 emitters over the entire time period have been China and The United States, both exceding nearlly 4x or more the amount of every other country.

- Throughout the time period, China and India have increased there CO2 Emissions the most out of every other country.
- Throughout the time period, Former soviet republics have had the largest decrease in CO2 emission, The United Kingdom and Germany have also decreased there emissions a bit as well.
- Generally speaking, the larger the population, the more CO2 the country will be likely to emit.
- The larger the GDP, the more likely the country will have a high CO2 emission.
- The larger the Energy Consumption of a country, the larger the CO2 emission.
- A high or low Energy Intensity by GDP of Energy Intensity per capita isnt necesarilly predictive of a large CO2 emission, but generally speaking the lower it is the better (the more energy conserved means less CO2 emitted).

References:

Kaggle dataset link: https://www.kaggle.com/datasets/lobosi/c02-emission-by-countrys-grouth-and-population

Organizations:
1. Analytics vidya
2. Geek for geeks

Acknowledgement:

Tools:
1. Python Libraries like Pandas, Numpy, Matplotlib, Seaborn,Scipy,Plotly,Math
2. Jupyter Notebook