# THE CURIOSITY CUP 2024
## A Global SAS® Student Competition

**Developing a Framework to Analyze World Energy Consumption Using SAS® Enterprise Guide**

Ki Jeong, Niharika Bandlamudi, Nupoor Karnik and Gagana Uday Kumar-University of Texas San Antonio


## ABSTRACT

This report presents the correlation found for energy consumption and population growth around several countries using SAS® Enterprise Guide. It also explores renewable vs. nonrenewable consumption rates to identify potential trends. Two datasets were used for the analyses: Population Data and World Energy Consumption Data. The aim was to allude that there may now be other significant factors driving a country's energy consumption than just mere population growth.


## INTRODUCTION

OUR WORLD IS CONSTANTLY EVOLVING WITH NEW ADVANCEMENTS IN TECHNOLOGY, WHICH MEANS HIGHER ENERGY CONSUMPTION. ADDITIONALLY, ACCORDING TO THE UNITED NATIONS, THE WORLD POPULATION IS PROJECTED TO REACH 8.5 BILLION BY 2030 AND 9.7 BILLION BY 2050. IN THIS REPORT, WE EXPLORE THE RELATIONSHIP BETWEEN POPULATION AND ENERGY CONSUMPTION TO REVEAL INSIGHTS ACROSS VARIOUS COUNTRIES.

ALTHOUGH CORRELATION BETWEEN POPULATION GROWTH AND ENERGY CONSUMPTION WAS ANTICIPATED, THERE WERE NOTABLE EXCEPTIONS. FURTHERMORE, THE ENERGY CONSUMPTION TRENDS DIFFERED BASED ON THE TYPE OF ENERGY THE COUNTRIES CONSUMED, I.E. RENEWABLE VS. NON-RENEWABLE. THESE EXCEPTIONS PROVIDED CLUES ON HOW THE REST OF THE WORLD MAY EVENTUALLY TREND TOWARDS IN THE FUTURE.

## PROBLEM STATEMENT

It's generally understood that energy consumption increases with population growth. But things may not be as simple as they seem as socially-conscious "green" movements arise from developed countries. We hoped that by visualizing the correlation between population and energy consumption, valuable information may be revealed that could tell a more insightful story behind the different countries and of the world.

## DATA PREPARATION

### DATASETS

In this project, we are using two datasets sourced from Kaggle and the World Bank website. The World Energy Consumption data, obtained from Kaggle, encompasses information on the country of energy consumption, the corresponding year, region, type of energy consumed and measured in exajoules. Simultaneously, the World Population data from the World Bank website, includes details such as country name, country code, year, Indicator Name, Indicator Code, and population figures.

We have created a library named "Kingg" and have established an autoexec process flow. We imported the files and performed comprehensive data profiling using Characterize Data Tasks to gain insights and summarize key attributes.

**PREPROCESSING DATA**

The data is cleaned and refined. First, a new table labeled 'population' is made selectively extracting essential columns and rectifying misspelled country names. Next, recognizing the enormity of the world energy consumption dataset, the top 20 most populous countries from 1995 to 2021 are meticulously selected, forming a more manageable table. Additionally, instances of '0' values are replaced with null entries to prevent their inclusion in average calculations. Moreover, country names are standardized across both datasets to ensure consistency and maintain unique key variables. Notably, 'Nigeria,' 'Ethiopia,' and 'Congo, Dem. Rep.' are omitted from the 'population' table due to their absence in the world energy consumption data, resulting in a refined dataset comprising 17 countries.

## VALIDATION

The summary statistics output for the POPULATION table and ENERGY_CONSUMPTION_TOP20 table validates the successful cleaning of the data, demonstrating effective grouping by years.
Utilizing functions such as CREATE, UPDATE, and DELETE, we have efficiently managed the data. You can find the examples for each function in appendix (fig2.1,2.2,2.3) to illustrate their implementation.

## ADVANCE PROC SQL AND VISUALIZATION

ENERGY_CONSUMPTION_TOP20 table is updated to include the new variable “Energy_source.” According to the energy type, each source is categorized to be either “Renewable” or “Non-Renewable.” This updated table is merged with the Population table via inner join on the common variables “Country” and “Year”. The results are grouped by country and ordered in descending order. Following code is used to obtain desired table:

```
Proc Sql;
 CREATE TABLE KINGG.MERGED_POPULATION_ENERGY AS
 Select*
 From KINGG.ENERGY_CONSUMPTION_TOP20 AS e
 Inner Join KINGG.POPULATION AS p
 On e.Country= p.Country_Name and e.Year = p.Year
 GROUP BY COUNTRY
 ORDER BY ENERGY DESC;
 Quit;
```

Merged_Population_Energy table is used to create pie charts grouped by year using the pie chart task to see the gradual increase of energy consumption from 1995 to 2021. The chart split depending on the Energy_source, “Renewable” or “Nonrenewable” as you see below.

As seen in fig:1.0 and fig:1.1, the total renewable energy increase in exajoules from 14.77 to 53.98 can be seen from 1995 to 2021. So, it can be said that we are moving towards a greener and cleaner future.

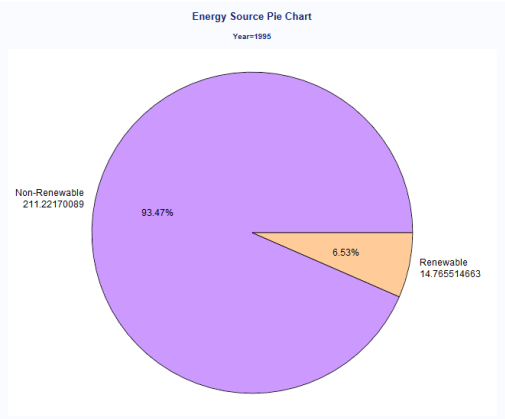


**fig: 1.0**

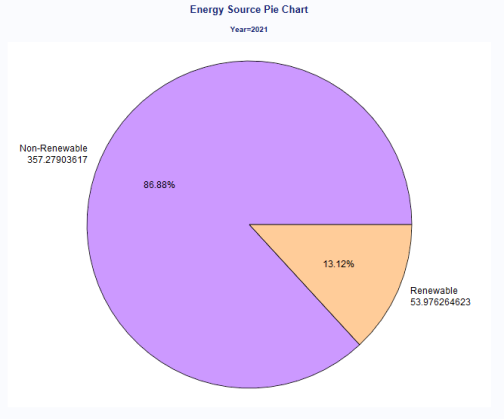


**fig:1.1**

To show the correlations between population and energy consumption, we have used Dual Axis Time Series Graph like below:

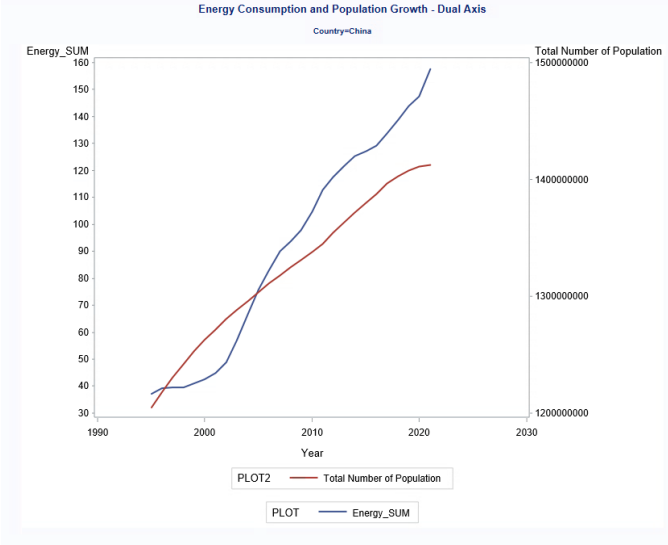


**fig:1.2**

China graph as an example.
From the dual axis time series graphs it is seen that the Population and Energy Consumption are correlated for 13 out of 17 countries which indicates that when we see change in population a similar trend is observed in energy consumption. This accounted for 76.47% of data.

## DATA ANALYSIS AND RESULTS

### CUSTOMIZED SUMMARY STATISTICS

To know which country consumes the highest amount of energy for each type we have used summary table:

Total Energy Consumption by Country and Type

| | Biofuel | Coal | Hydro-electric | Natural gas | Nuclear energy | Oil | Other renewables | Solar | Wind | Energy Consumption per Country |
|---|---|---|---|---|---|---|---|---|---|---|
| Bangladesh | | 1.28 | 0.21 | 17.01 | | 5.83 | 0.00 | 0.03 | 0.00 | 24.35 |
| Brazil | 13.17 | 15.98 | 94.07 | 23.07 | 3.08 | 113.35 | 8.52 | 0.37 | 3.47 | 275.08 |
| China | 1.52 | 1617.91 | 174.45 | 119.18 | 28.77 | 461.27 | 10.99 | 11.90 | 30.71 | 2456.70 |
| Germany | 2.09 | 88.38 | 5.54 | 83.03 | 34.25 | 138.10 | 8.36 | 4.25 | 12.50 | 376.50 |
| India | 0.98 | 314.40 | 29.71 | 39.82 | 6.39 | 172.26 | 4.15 | 2.48 | 5.73 | 575.92 |
| Indonesia | 1.33 | 40.82 | 3.65 | 37.32 | | 70.92 | 2.75 | 0.01 | 0.02 | 156.81 |
| Iran | | 1.59 | 3.27 | 123.81 | 0.46 | 85.67 | 0.00 | 0.02 | 0.06 | 214.89 |
| Japan | | 122.34 | 22.21 | 91.91 | 51.48 | 260.13 | 6.70 | 4.72 | 0.83 | 560.32 |
| Mexico | 0.12 | 11.37 | 8.16 | 58.47 | 2.63 | 98.35 | 1.91 | 0.32 | 1.14 | 182.45 |
| Pakistan | | 5.95 | 7.46 | 29.65 | 0.96 | 23.11 | 0.11 | 0.07 | 0.15 | 67.46 |
| Philippines | | 8.87 | 2.26 | 2.45 | | 19.33 | 2.67 | 0.07 | 0.08 | 35.73 |
| Russian Federation | | 108.90 | 47.34 | 394.95 | 41.91 | 159.67 | 0.10 | 0.07 | 0.05 | 752.98 |
| Thailand | 0.86 | 15.19 | 1.67 | 34.58 | | 53.56 | 1.33 | 0.33 | 0.14 | 107.66 |
| Turkey | | 32.91 | 12.88 | 30.16 | | 40.69 | 0.94 | 0.43 | 1.68 | 119.69 |
| United States | 20.71 | 510.06 | 76.27 | 652.38 | 211.50 | 991.31 | 22.11 | 7.13 | 28.68 | 2520.14 |
| Vietnam | | 18.98 | 9.97 | 5.79 | | 17.68 | 0.02 | 0.40 | 0.05 | 52.88 |
| Energy Consumption per Type | 40.78 | 2914.93 | 499.10 | 1743.58 | 381.42 | 2711.22 | 70.66 | 32.58 | 85.29 | 8479.55 |

**Table-1.0**

In the above Table-1.0 we see that "China" and "US" are the highest energy consumers in the world from (1995 to 2021).
The most energy type used is "Coal" and "Oil" from 1995 to 2021.
To compare the average of the renewable energy consumption by top countries in all years with the entire data of renewable energy to see which country surpassed this average and in which year using an inline view. The same process is followed for non-renewable energy. (fig:2.4 in appendix for SAS® code)
To understand energy consumed per person (renewable/non-renewable) by country we have used a new table named 'Total_energy_country_with_source 'to see the per person energy consumed in each year of each Country (converted from Exajoules to Gigajoules) by energy source combining the 'Total_energy_country table' and 'Population table'.
Following are the line graphs to visualize the results of above Table-1.0:

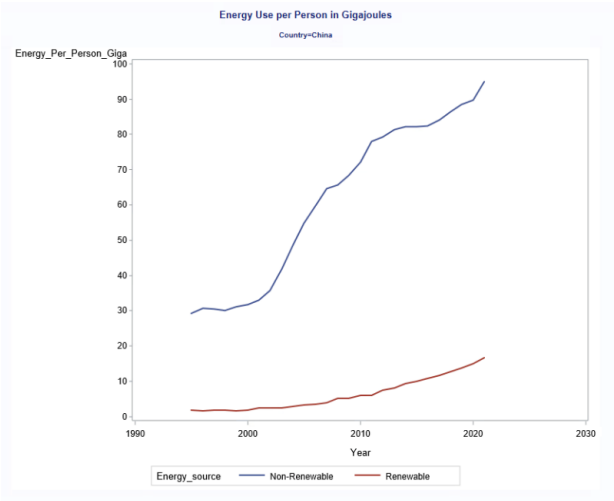


**fig:1.3**

13 out of 17 countries shared this kind of pattern as shown in fig:1.3: Increasing non-renewable energy use, while very slow to no-growth in renewable energy use. This accounted for 76.47% of data. Figures were expressed in gigajoules.

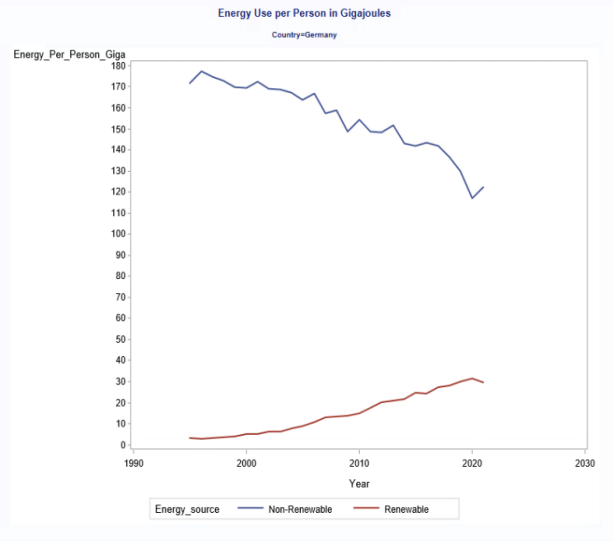


**fig:1.4**

4 out of 17 countries shared this kind of pattern as shown in fig:1.4: Decreasing non-renewable energy use, while increasing renewable energy use. This accounted for 23.53% of data. Figures were expressed in gigajoules.

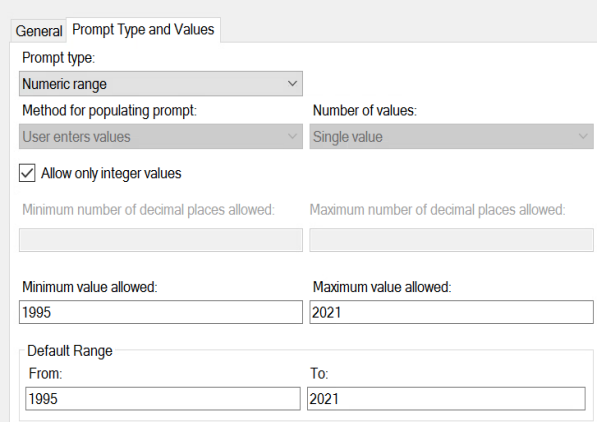


**fig:1.5- Prompts**

We have created prompts for the year to select the year range and have given numeric range as type of prompt and assigned the default range. We have used this prompt in TOTAL_ENERGY_COUNTRY_WITH_SOURCE. It gives the leverage to view the data. Output from the query shows Year_Range_Prompt is working correctly by selecting only the years indicated.

## CONCLUSION

In examining the correlation between population and energy consumption, we find a strong correlation of 76.47%, indicating a positive relationship between the two variables. This suggests that as population increases, energy consumption also tends to increase. Moreover, in understanding the energy consumption per person, we observe that 76.47% of the cases demonstrate a pattern of increasing non-renewable energy usage alongside constant renewable energy usage. Conversely, in 23.53% of cases, there is a trend of declining non-renewable energy usage while renewable energy consumption rises. These findings shed light on the dynamics of energy consumption patterns across different countries, highlighting both the prevailing correlation between population and energy usage, as well as the varying trajectories of renewable and non-renewable energy consumption.

## RECOMMENDATIONS

Future researchers are recommended that for a comprehensive analysis, it is beneficial to expand the scope to include all countries in both datasets, rather than limiting the investigation to just 17 countries. Additionally, introducing a new variable for clean energy is essential, as renewable energy sources may not always align with environmental sustainability goals. Considering the climate is also crucial, as it influences energy consumption patterns. Moreover, factors such as population density, with variations between sparse and densely populated countries, are expected to play a significant role.

Furthermore, a country's reliance on industry or agriculture is likely to impact its energy consumption trends. These considerations underscore the need for a multifaceted approach to understand the complexities of energy consumption patterns globally.

## REFERENCES

*Population, total*. World Bank Open Data. (2022). https://data.worldbank.org/indicator/SP.POP.TOTL

Prasad, N. (2022, December 10). *World Energy Consumption*. Kaggle. https://www.kaggle.com/datasets/nirmalprasad/world-energy-consumption

*SAS® Enterprise Guide® 8.3: User's Guide*. SAS® help center. (2024, February 8). https://documentation.sas.com/doc/en/egcdc/8.3/egdoccdc/egug/titlepage.htm

### CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Gagana Uday Kumar gagana.udaykumar@my.utsa.edu



## APPENDIX A

DESCRIBE DATA statement to explore the variables and their attributes:

```
PROC SQL;
DESCRIBE TABLE KINGG.POPULATION_PRECLEAN;
DESCRIBE TABLE KINGG.WORLD_ENERGY_CONSUMPTION;
QUIT;
```

The results of these queries are below:

```
CREATE TABLE KINGG.POPULATION_PRECLEAN( BUFSIZE=65536 )
  (
   COUNTRY_NAME CHAR(18) FORMAT=$CHAR18. INFORMAT=$CHAR18. LABEL='COUNTRY
NAME',
   COUNTRY_CODE CHAR(3) FORMAT=$CHAR3. INFORMAT=$CHAR3. LABEL='COUNTRY CODE',
   YEAR NUM FORMAT=BEST12. INFORMAT=BEST12.,
   INDICATOR_NAME CHAR(17) FORMAT=$CHAR17. INFORMAT=$CHAR17.,
   INDICATOR_CODE CHAR(11) FORMAT=$CHAR11. INFORMAT=$CHAR11. LABEL='INDICATOR
CODE',
   TOTAL_POPULATION NUM FORMAT=BEST12. INFORMAT=BEST12. LABEL='TOTAL NUMBER
```

```
OF POPULATION'  );
CREATE TABLE KINGG.WORLD_ENERGY_CONSUMPTION( BUFSIZE=65536 )
  (
   COUNTRY CHAR(24) FORMAT=$CHAR24. INFORMAT=$CHAR24.,
   YEAR NUM FORMAT=BEST12. INFORMAT=BEST12.,
   REGION CHAR(18) FORMAT=$CHAR18. INFORMAT=$CHAR18.,
   TYPE CHAR(16) FORMAT=$CHAR16. INFORMAT=$CHAR16.,
   ENERGY NUM FORMAT=BEST12.11 INFORMAT=BEST12.11 LABEL='ENERGY CONSUMPTION
(IN EXAJOULES)'  );
```

**fig 2.1**

```
/*SELECTING THE TOP 20 MOST PUPULATED COUNTRIES
AND CREATED A NEW TABLE NAMED
ENERGY_CONSUMPTION_TOP20 AS
WORLD_ENERGY_CONSUMPTION WAS A HUGE DATA */

PROC SQL;
    CREATE TABLE KINGG.ENERGY_CONSUMPTION_TOP20 AS
    SELECT Country, Year, Region, Type, Energy
    FROM KINGG.WORLD_ENERGY_CONSUMPTION
    WHERE Country IN ('India', 'China', 'United States', 'Indonesia',
'Pakistan', 'Nigeria', 'Brazil', 'Bangladesh', 'Russian Federation',
'Mexico', 'Japan', 'Ethiopia', 'Philippines', 'Egypt, Arab Rep.',
'Congo, Dem. Rep.', 'Vietnam', 'Iran', 'Turkey', 'Germany', 'Thailand');
QUIT;
```

**fig 2.2**

```
/* CHANGING THE COUNTRY NAMES TO MAINTAIN
THE UNIQUE KEY VARIABLES IN POPULATION TABLE*/

PROC SQL;
    UPDATE KINGG.POPULATION
    SET Country_Name =
       CASE
          WHEN Country_Name = 'Egypt, Arab Rep.' THEN 'Egypt'
          WHEN Country_Name = 'Viet Nam' THEN 'Vietnam'
          WHEN Country_Name = 'Iran, Islamic Rep.' THEN 'Iran'
          WHEN Country_Name = 'Turkiye' THEN 'Turkey'
          ELSE Country_Name
       END;
QUIT;
```

**fig 2.3**

```
/*REMOVING THE BELOW COUNTRIES FROM THE POPULATION
TABLE AS THEIR ENERGY CONSUMPTION DATA WERE NOT FOUND*/

PROC SQL;
    DELETE FROM KINGG.POPULATION
    WHERE COUNTRY_NAME = 'Nigeria';
QUIT;
```

**fig 2.4**

```
PROC SQL;
SELECT Country, Year, type, Energy
from KINGG.ENERGY_CONSUMPTION_TOP20
where Energy_source='Non-Renewable' and Energy > (select avg(Energy)
                from KINGG.ENERGY_CONSUMPTION_TOP20
                where Energy_source='Non-Renewable')
order by Energy Desc;
quit;
```