01_data_exploration

November 30, 2024

```
Looking for data file at:
/Users/katejohnson/Documents/Other/Northeastern/CS6140/Course
Project/cs6140-course-project/processed_data/final_processed_data.csv

Dataset Overview:
================================================================================

Shape: (613, 8)

Features:
- Year: float64 (Missing: 0)
- Month: float64 (Missing: 0)
- Hydroelectric Power: float64 (Missing: 0)
- Solar Energy: float64 (Missing: 0)
- Wind Energy: float64 (Missing: 0)
- Geothermal Energy: float64 (Missing: 0)
- Biomass Energy: float64 (Missing: 0)
- Total Renewable Energy: float64 (Missing: 0)

Current working directory:
/Users/katejohnson/Documents/Other/Northeastern/CS6140/Course
Project/cs6140-course-project/notebooks

Trying to load datasets…
Loading data from: /Users/katejohnson/Documents/Other/Northeastern/CS6140/Course
Project/cs6140-course-project/data
Checking global energy path:
/Users/katejohnson/Documents/Other/Northeastern/CS6140/Course
Project/cs6140-course-project/data/Global Energy Consumption & Renewable
Generation
Path exists: True

Datasets loaded successfully!
Loading data from: /Users/katejohnson/Documents/Other/Northeastern/CS6140/Course
Project/cs6140-course-project/data
Checking global energy path:
/Users/katejohnson/Documents/Other/Northeastern/CS6140/Course
Project/cs6140-course-project/data/Global Energy Consumption & Renewable
Generation
```

```
Path exists: True


Global Energy Consumption & Renewable Generation Datasets
================================================================================

Dataset: continent_consumption
Shape: (31, 12)

Columns:
- Year: int64 (Missing: 0)
- World: float64 (Missing: 0)
- OECD: float64 (Missing: 0)
- BRICS: float64 (Missing: 0)
- Europe: float64 (Missing: 0)
- North America: float64 (Missing: 0)
- Latin America: float64 (Missing: 0)
- Asia: float64 (Missing: 0)
- Pacific: float64 (Missing: 0)
- Africa: float64 (Missing: 0)
- Middle-East: float64 (Missing: 0)
- CIS: float64 (Missing: 0)
----------------------------------------

Dataset: country_consumption
Shape: (33, 45)

Columns:
- Year: float64 (Missing: 2)
- China: float64 (Missing: 2)
- United States: float64 (Missing: 2)
- Brazil: float64 (Missing: 2)
- Belgium: float64 (Missing: 2)
- Czechia: float64 (Missing: 2)
- France: float64 (Missing: 2)
- Germany: float64 (Missing: 2)
- Italy: float64 (Missing: 2)
- Netherlands: float64 (Missing: 2)
- Poland: float64 (Missing: 2)
- Portugal: float64 (Missing: 2)
- Romania: float64 (Missing: 2)
- Spain: float64 (Missing: 2)
- Sweden: float64 (Missing: 2)
- United Kingdom: float64 (Missing: 2)
- Norway: float64 (Missing: 2)
- Turkey: float64 (Missing: 2)
- Kazakhstan: float64 (Missing: 2)
- Russia: float64 (Missing: 2)
```

- Ukraine: float64 (Missing: 2)
- Uzbekistan: float64 (Missing: 2)
- Argentina: float64 (Missing: 2)
- Canada: float64 (Missing: 2)
- Chile: float64 (Missing: 2)
- Colombia: float64 (Missing: 2)
- Mexico: float64 (Missing: 2)
- Venezuela: float64 (Missing: 2)
- Indonesia: float64 (Missing: 2)
- Japan: float64 (Missing: 2)
- Malaysia: float64 (Missing: 2)
- South Korea: float64 (Missing: 2)
- Taiwan: float64 (Missing: 2)
- Thailand: float64 (Missing: 2)
- India: float64 (Missing: 2)
- Australia: float64 (Missing: 2)
- New Zealand: float64 (Missing: 2)
- Algeria: float64 (Missing: 2)
- Egypt: float64 (Missing: 2)
- Nigeria: float64 (Missing: 2)
- South Africa: float64 (Missing: 2)
- Iran: float64 (Missing: 2)
- Kuwait: float64 (Missing: 2)
- Saudi Arabia: float64 (Missing: 2)
- United Arab Emirates: float64 (Missing: 2)

Dataset: renewable_gen
Shape: (28, 5)

Columns:
- Year: int64 (Missing: 0)
- Hydro(TWh): float64 (Missing: 0)
- Biofuel(TWh): float64 (Missing: 0)
- Solar PV (TWh): float64 (Missing: 0)
- Geothermal (TWh): float64 (Missing: 0)

Dataset: nonrenewable_gen
Shape: (8, 2)

Columns:
- Mode of Generation: object (Missing: 0)
- Contribution (TWh): float64 (Missing: 0)

Worldwide Renewable Energy Datasets
==

```
Dataset: renewable_share
Shape: (5603, 4)

Columns:
- Entity: object (Missing: 0)
- Code: object (Missing: 1311)
- Year: int64 (Missing: 0)
- Renewables (% equivalent primary energy): float64 (Missing: 0)
----------------------------------------

Dataset: renewable_consumption
Shape: (5610, 7)

Columns:
- Entity: object (Missing: 0)
- Code: object (Missing: 1311)
- Year: int64 (Missing: 0)
- Geo Biomass Other - TWh: float64 (Missing: 144)
- Solar Generation - TWh: float64 (Missing: 168)
- Wind Generation - TWh: float64 (Missing: 165)
- Hydro Generation - TWh: float64 (Missing: 7)
----------------------------------------

Dataset: hydro_consumption
Shape: (8840, 4)

Columns:
- Entity: object (Missing: 0)
- Code: object (Missing: 1555)
- Year: int64 (Missing: 0)
- Electricity from hydro (TWh): float64 (Missing: 0)
----------------------------------------

Dataset: wind_generation
Shape: (8676, 4)

Columns:
- Entity: object (Missing: 0)
- Code: object (Missing: 1459)
- Year: int64 (Missing: 0)
- Electricity from wind (TWh): float64 (Missing: 0)
----------------------------------------

Dataset: solar_consumption
Shape: (8683, 4)

Columns:
```

```
- Entity: object (Missing: 0)
- Code: object (Missing: 1456)
- Year: int64 (Missing: 0)
- Electricity from solar (TWh): float64 (Missing: 0)
----------------------------------------

Weather Conditions Dataset
================================================================================
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 196776 entries, 0 to 196775
Data columns (total 17 columns):
 #   Column                   Non-Null Count   Dtype
---  ------                   --------------   -----
 0   Time                     196776 non-null  object
 1   Energy delta[Wh]         196776 non-null  int64
 2   GHI                      196776 non-null  float64
 3   temp                     196776 non-null  float64
 4   pressure                 196776 non-null  int64
 5   humidity                 196776 non-null  int64
 6   wind_speed               196776 non-null  float64
 7   rain_1h                  196776 non-null  float64
 8   snow_1h                  196776 non-null  float64
 9   clouds_all               196776 non-null  int64
 10  isSun                    196776 non-null  int64
 11  sunlightTime             196776 non-null  int64
 12  dayLength                196776 non-null  int64
 13  SunlightTime/daylength   196776 non-null  float64
 14  weather_type             196776 non-null  int64
 15  hour                     196776 non-null  int64
 16  month                    196776 non-null  int64
dtypes: float64(6), int64(10), object(1)
memory usage: 25.5+ MB
None

US Renewable Energy Dataset
================================================================================
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 3065 entries, 0 to 3064
Data columns (total 17 columns):
 #   Column                               Non-Null Count  Dtype
---  ------                               --------------  -----
 0   Year                                 3065 non-null   int64
 1   Month                                3065 non-null   int64
 2   Sector                               3065 non-null   object
 3   Hydroelectric Power                  3065 non-null   float64
 4   Geothermal Energy                    3065 non-null   float64
 5   Solar Energy                         3065 non-null   float64
```

```
 6   Wind Energy                          3065 non-null   float64
 7   Wood Energy                          3065 non-null   float64
 8   Waste Energy                         3065 non-null   float64
 9   Fuel Ethanol, Excluding Denaturant   3065 non-null   float64
 10  Biomass Losses and Co-products       3065 non-null   float64
 11  Biomass Energy                       3065 non-null   float64
 12  Total Renewable Energy               3065 non-null   float64
 13  Renewable Diesel Fuel                3065 non-null   float64
 14  Other Biofuels                       3065 non-null   float64
 15  Conventional Hydroelectric Power     3065 non-null   float64
 16  Biodiesel                            3065 non-null   float64
dtypes: float64(14), int64(2), object(1)
memory usage: 407.2+ KB

None


Global Energy Consumption & Renewable Generation Datasets
================================================================================

Dataset: continent_consumption
Shape: (31, 12)

Columns:
- Year: int64 (Missing: 0)
- World: float64 (Missing: 0)
- OECD: float64 (Missing: 0)
- BRICS: float64 (Missing: 0)
- Europe: float64 (Missing: 0)
- North America: float64 (Missing: 0)
- Latin America: float64 (Missing: 0)
- Asia: float64 (Missing: 0)
- Pacific: float64 (Missing: 0)
- Africa: float64 (Missing: 0)
- Middle-East: float64 (Missing: 0)
- CIS: float64 (Missing: 0)
----------------------------------------

Dataset: country_consumption
Shape: (33, 45)

Columns:
- Year: float64 (Missing: 2)
- China: float64 (Missing: 2)
- United States: float64 (Missing: 2)
- Brazil: float64 (Missing: 2)
- Belgium: float64 (Missing: 2)
- Czechia: float64 (Missing: 2)
- France: float64 (Missing: 2)
```

- Germany: float64 (Missing: 2)
- Italy: float64 (Missing: 2)
- Netherlands: float64 (Missing: 2)
- Poland: float64 (Missing: 2)
- Portugal: float64 (Missing: 2)
- Romania: float64 (Missing: 2)
- Spain: float64 (Missing: 2)
- Sweden: float64 (Missing: 2)
- United Kingdom: float64 (Missing: 2)
- Norway: float64 (Missing: 2)
- Turkey: float64 (Missing: 2)
- Kazakhstan: float64 (Missing: 2)
- Russia: float64 (Missing: 2)
- Ukraine: float64 (Missing: 2)
- Uzbekistan: float64 (Missing: 2)
- Argentina: float64 (Missing: 2)
- Canada: float64 (Missing: 2)
- Chile: float64 (Missing: 2)
- Colombia: float64 (Missing: 2)
- Mexico: float64 (Missing: 2)
- Venezuela: float64 (Missing: 2)
- Indonesia: float64 (Missing: 2)
- Japan: float64 (Missing: 2)
- Malaysia: float64 (Missing: 2)
- South Korea: float64 (Missing: 2)
- Taiwan: float64 (Missing: 2)
- Thailand: float64 (Missing: 2)
- India: float64 (Missing: 2)
- Australia: float64 (Missing: 2)
- New Zealand: float64 (Missing: 2)
- Algeria: float64 (Missing: 2)
- Egypt: float64 (Missing: 2)
- Nigeria: float64 (Missing: 2)
- South Africa: float64 (Missing: 2)
- Iran: float64 (Missing: 2)
- Kuwait: float64 (Missing: 2)
- Saudi Arabia: float64 (Missing: 2)
- United Arab Emirates: float64 (Missing: 2)

Dataset: renewable_gen
Shape: (28, 5)

Columns:
- Year: int64 (Missing: 0)
- Hydro(TWh): float64 (Missing: 0)
- Biofuel(TWh): float64 (Missing: 0)
- Solar PV (TWh): float64 (Missing: 0)

```
- Geothermal (TWh): float64 (Missing: 0)
----------------------------------------

Dataset: nonrenewable_gen
Shape: (8, 2)

Columns:
- Mode of Generation: object (Missing: 0)
- Contribution (TWh): float64 (Missing: 0)
----------------------------------------

Worldwide Renewable Energy Datasets
================================================================================

Dataset: renewable_share
Shape: (5603, 4)

Columns:
- Entity: object (Missing: 0)
- Code: object (Missing: 1311)
- Year: int64 (Missing: 0)
- Renewables (% equivalent primary energy): float64 (Missing: 0)
----------------------------------------

Dataset: renewable_consumption
Shape: (5610, 7)

Columns:
- Entity: object (Missing: 0)
- Code: object (Missing: 1311)
- Year: int64 (Missing: 0)
- Geo Biomass Other - TWh: float64 (Missing: 144)
- Solar Generation - TWh: float64 (Missing: 168)
- Wind Generation - TWh: float64 (Missing: 165)
- Hydro Generation - TWh: float64 (Missing: 7)
----------------------------------------

Dataset: hydro_consumption
Shape: (8840, 4)

Columns:
- Entity: object (Missing: 0)
- Code: object (Missing: 1555)
- Year: int64 (Missing: 0)
- Electricity from hydro (TWh): float64 (Missing: 0)
----------------------------------------

Dataset: wind_generation
```

```
Shape: (8676, 4)

Columns:
- Entity: object (Missing: 0)
- Code: object (Missing: 1459)
- Year: int64 (Missing: 0)
- Electricity from wind (TWh): float64 (Missing: 0)
----------------------------------------

Dataset: solar_consumption
Shape: (8683, 4)

Columns:
- Entity: object (Missing: 0)
- Code: object (Missing: 1456)
- Year: int64 (Missing: 0)
- Electricity from solar (TWh): float64 (Missing: 0)
----------------------------------------

Weather Conditions Dataset
================================================================================
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 196776 entries, 0 to 196775
Data columns (total 17 columns):
 #   Column                  Non-Null Count   Dtype
---  ------                  --------------   -----
 0   Time                    196776 non-null  object
 1   Energy delta[Wh]        196776 non-null  int64
 2   GHI                     196776 non-null  float64
 3   temp                    196776 non-null  float64
 4   pressure                196776 non-null  int64
 5   humidity                196776 non-null  int64
 6   wind_speed              196776 non-null  float64
 7   rain_1h                 196776 non-null  float64
 8   snow_1h                 196776 non-null  float64
 9   clouds_all              196776 non-null  int64
 10  isSun                   196776 non-null  int64
 11  sunlightTime            196776 non-null  int64
 12  dayLength               196776 non-null  int64
 13  SunlightTime/daylength  196776 non-null  float64
 14  weather_type            196776 non-null  int64
 15  hour                    196776 non-null  int64
 16  month                   196776 non-null  int64
dtypes: float64(6), int64(10), object(1)
memory usage: 25.5+ MB

None
```

```
US Renewable Energy Dataset
================================================================================
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 3065 entries, 0 to 3064
Data columns (total 17 columns):
 #   Column                              Non-Null Count  Dtype
---  ------                              --------------  -----
 0   Year                                3065 non-null   int64
 1   Month                               3065 non-null   int64
 2   Sector                              3065 non-null   object
 3   Hydroelectric Power                 3065 non-null   float64
 4   Geothermal Energy                   3065 non-null   float64
 5   Solar Energy                        3065 non-null   float64
 6   Wind Energy                         3065 non-null   float64
 7   Wood Energy                         3065 non-null   float64
 8   Waste Energy                        3065 non-null   float64
 9   Fuel Ethanol, Excluding Denaturant  3065 non-null   float64
 10  Biomass Losses and Co-products      3065 non-null   float64
 11  Biomass Energy                      3065 non-null   float64
 12  Total Renewable Energy              3065 non-null   float64
 13  Renewable Diesel Fuel               3065 non-null   float64
 14  Other Biofuels                      3065 non-null   float64
 15  Conventional Hydroelectric Power    3065 non-null   float64
 16  Biodiesel                           3065 non-null   float64
dtypes: float64(14), int64(2), object(1)
memory usage: 407.2+ KB

None


Global Energy Data Quality Assessment
================================================================================

Dataset: continent_consumption

Duplicate Rows: 0

Numerical Columns Statistics:
          Year      World      OECD     BRICS    Europe  North America  \
count    31.00      31.00     31.00     31.00     31.00          31.00
mean   2005.00  132792.47  60396.47  41128.93  21487.74       28226.76
std       9.09   22724.12   3480.62  13849.97    899.17        1548.24
min    1990.00  101855.54  52602.49  25993.05  19643.07       24667.23
25%    1997.50  111176.98  58719.87  27504.95  20875.85       27435.17
50%    2005.00  133582.18  61545.96  38169.66  21480.61       28598.17
75%    2012.50  154853.45  62360.06  55521.62  21951.62       29295.97
max    2020.00  167553.41  64883.77  63255.57  23108.81       30424.08

       Latin America      Asia  Pacific   Africa  Middle-East       CIS
```

count	31.00	31.00	31.00	31.00	31.00	31.00
mean	7897.15	45402.02	1563.30	6851.95	5984.20	11823.96
std	1537.72	15511.85	205.51	1742.66	2245.55	1410.09
min	5373.06	24574.19	1186.26	4407.77	2581.86	10152.99
25%	6687.25	31383.56	1424.68	5355.62	4070.50	11001.98
50%	8059.59	43693.91	1570.05	6652.36	5675.44	11606.74
75%	9391.22	60760.94	1756.13	8367.78	8007.26	12083.57
max	9978.54	69582.29	1802.65	9641.27	9455.19	16049.40

Dataset: country_consumption

Missing Values:

Year	2
China	2
United States	2
Brazil	2
Belgium	2
Czechia	2
France	2
Germany	2
Italy	2
Netherlands	2
Poland	2
Portugal	2
Romania	2
Spain	2
Sweden	2
United Kingdom	2
Norway	2
Turkey	2
Kazakhstan	2
Russia	2
Ukraine	2
Uzbekistan	2
Argentina	2
Canada	2
Chile	2
Colombia	2
Mexico	2
Venezuela	2
Indonesia	2
Japan	2
Malaysia	2
South Korea	2
Taiwan	2
Thailand	2
India	2

```
Australia              2
New Zealand            2
Algeria                2
Egypt                  2
Nigeria                2
South Africa           2
Iran                   2
Kuwait                 2
Saudi Arabia           2
United Arab Emirates   2
dtype: int64

Duplicate Rows: 1

Numerical Columns Statistics:
          Year    China  United States  Brazil  Belgium  Czechia  France  \
count    31.00    31.00          31.00   31.00    31.00    31.00   31.00
mean   2005.00  1923.32        2167.45  223.45    54.90    43.26  251.19
std       9.09   898.86         114.08   55.46     3.03     2.19   13.64
min    1990.00   848.00        1910.00  141.00    48.00    39.00  217.00
25%    1997.50  1076.50        2119.00  181.00    53.00    42.00  243.50
50%    2005.00  1782.00        2191.00  216.00    56.00    43.00  252.00
75%    2012.50  2866.50        2246.00  284.00    57.00    45.00  260.50
max    2020.00  3381.00        2338.00  303.00    60.00    50.00  273.00

       Germany   Italy  Netherlands  …  Australia  New Zealand  Algeria  \
count     31.0   31.00        31.00  …      31.00        31.00    31.00
mean     327.9  162.90        74.87  …     112.65        17.61    37.26
std       18.4   14.02         3.98  …      14.99         2.25    13.75
min      275.0  137.00        67.00  …      85.00        14.00    22.00
25%      313.0  150.50        72.00  …     102.50        16.00    24.50
50%      335.0  162.00        75.00  …     113.00        17.00    32.00
75%      340.0  173.00        77.50  …     126.50        19.00    48.00
max      351.0  187.00        83.00  …     129.00        21.00    65.00

       Egypt  Nigeria  South Africa    Iran  Kuwait  Saudi Arabia  \
count  31.00    31.00         31.00   31.00   31.00         31.00
mean   60.94   108.97        118.19  169.06   23.16        138.39
std    21.91    31.86         16.72   64.86    9.04         53.97
min    33.00    66.00         88.00   69.00    3.00         58.00
25%    40.50    79.50        106.00  110.00   16.00         91.00
50%    62.00   105.00        120.00  173.00   25.00        123.00
75%    78.50   141.50        132.50  220.00   29.00        188.50
max    97.00   160.00        144.00  269.00   38.00        219.00

       United Arab Emirates
count                 31.00
mean                  49.06
```

```
std                   20.97
min                   20.00
25%                   31.00
50%                   44.00
75%                   66.00
max                   83.00

[8 rows x 45 columns]
----------------------------------------

Dataset: renewable_gen

Duplicate Rows: 0

Numerical Columns Statistics:
          Year  Hydro(TWh)  Biofuel(TWh)  Solar PV (TWh)  Geothermal (TWh)
count    28.00       28.00         28.00           28.00             28.00
mean   2003.50     2974.17        245.03           57.43             57.01
std       8.23      595.94        329.28          113.34             14.85
min    1990.00     2191.67          3.88            0.09             36.42
25%    1996.75     2598.63         11.42            0.26             42.33
50%    2003.50     2718.72         74.33            2.34             55.30
75%    2010.25     3298.90        365.04           40.10             68.40
max    2017.00     4197.29       1127.31          443.55             85.34
----------------------------------------

Dataset: nonrenewable_gen

Duplicate Rows: 0

Numerical Columns Statistics:
       Contribution (TWh)
count                8.00
mean              4862.04
std               6852.38
min                 36.02
25%                104.04
50%               1738.95
75%               6877.95
max              19448.16
----------------------------------------

Worldwide Renewable Data Quality Assessment
================================================================================

Dataset: renewable_share

Missing Values:
```

```
Code    1311
dtype: int64

Duplicate Rows: 0

Numerical Columns Statistics:
          Year  Renewables (% equivalent primary energy)
count  5603.00                                   5603.00
mean   1993.80                                     10.74
std      16.28                                     12.92
min    1965.00                                      0.00
25%    1980.00                                      1.98
50%    1994.00                                      6.52
75%    2008.00                                     14.10
max    2021.00                                     86.87
----------------------------------------

Dataset: renewable_consumption

Missing Values:
Code                        1311
Geo Biomass Other - TWh      144
Solar Generation - TWh       168
Wind Generation - TWh        165
Hydro Generation - TWh         7
dtype: int64

Duplicate Rows: 0

Numerical Columns Statistics:
          Year  Geo Biomass Other - TWh  Solar Generation - TWh  \
count  5610.00                  5466.00                 5442.00
mean   1993.83                    13.46                    5.48
std      16.30                    47.64                   39.90
min    1965.00                     0.00                    0.00
25%    1980.00                     0.00                    0.00
50%    1994.00                     0.23                    0.00
75%    2008.00                     4.27                    0.02
max    2021.00                   762.78                 1032.50

       Wind Generation - TWh  Hydro Generation - TWh
count                5445.00                 5603.00
mean                   15.03                  147.89
std                    84.73                  390.19
min                     0.00                    0.00
25%                     0.00                    1.37
50%                     0.00                   10.69
75%                     0.28                   65.84
```

```
max                  1861.94                 4345.99
----------------------------------------

Dataset: hydro_consumption

Missing Values:
Code    1555
dtype: int64

Duplicate Rows: 0

Numerical Columns Statistics:
          Year  Electricity from hydro (TWh)
count  8840.00                       8840.00
mean   1999.89                        116.58
std      15.75                        360.23
min    1965.00                          0.00
25%    1988.00                          0.09
50%    2004.00                          3.53
75%    2013.00                         30.07
max    2022.00                       4340.61
----------------------------------------

Dataset: wind_generation

Missing Values:
Code    1459
dtype: int64

Duplicate Rows: 0

Numerical Columns Statistics:
          Year  Electricity from wind (TWh)
count  8676.00                      8676.00
mean   2000.34                        14.57
std      15.51                        86.39
min    1965.00                         0.00
25%    1990.00                         0.00
50%    2004.00                         0.00
75%    2013.00                         0.06
max    2022.00                      1848.26
----------------------------------------

Dataset: solar_consumption

Missing Values:
Code    1456
dtype: int64
```

```
Duplicate Rows: 0

Numerical Columns Statistics:
          Year  Electricity from solar (TWh)
count  8683.00                       8683.00
mean   2000.38                          5.28
std      15.50                         40.10
min    1965.00                          0.00
25%    1990.00                          0.00
50%    2004.00                          0.00
75%    2013.00                          0.01
max    2022.00                       1040.50
----------------------------------------

Weather Data Quality Assessment
================================================================================

       Energy delta[Wh]            GHI           temp       pressure  \
count     196776.000000  196776.000000  196776.000000  196776.000000
mean         573.008228      32.596538       9.790521    1015.292780
std         1044.824047      52.172018       7.995428       9.585773
min            0.000000       0.000000     -16.600000     977.000000
25%            0.000000       0.000000       3.600000    1010.000000
50%            0.000000       1.600000       9.300000    1016.000000
75%          577.000000      46.800000      15.700000    1021.000000
max         5020.000000     229.200000      35.800000    1047.000000

            humidity     wind_speed        rain_1h        snow_1h  \
count  196776.000000  196776.000000  196776.000000  196776.000000
mean       79.810566       3.937746       0.066035       0.007148
std        15.604459       1.821694       0.278913       0.069710
min        22.000000       0.000000       0.000000       0.000000
25%        70.000000       2.600000       0.000000       0.000000
50%        84.000000       3.700000       0.000000       0.000000
75%        92.000000       5.000000       0.000000       0.000000
max       100.000000      14.300000       8.090000       2.820000

          clouds_all          isSun   sunlightTime      dayLength  \
count  196776.000000  196776.000000  196776.000000  196776.000000
mean       65.974387       0.519962     211.721094     748.644347
std        36.628593       0.499603     273.902186     194.870208
min         0.000000       0.000000       0.000000     450.000000
25%        34.000000       0.000000       0.000000     570.000000
50%        82.000000       1.000000      30.000000     765.000000
75%       100.000000       1.000000     390.000000     930.000000
max       100.000000       1.000000    1020.000000    1020.000000
```

	SunlightTime/daylength	weather_type	hour	month
count	196776.000000	196776.000000	196776.000000	196776.000000
mean	0.265187	3.198398	11.498902	6.298329
std	0.329023	1.289939	6.921887	3.376066
min	0.000000	1.000000	0.000000	1.000000
25%	0.000000	2.000000	5.000000	3.000000
50%	0.050000	4.000000	11.000000	6.000000
75%	0.530000	4.000000	17.000000	9.000000
max	1.000000	5.000000	23.000000	12.000000

US Data Quality Assessment

	Year	Month	Hydroelectric Power	Geothermal Energy
count	3065.000000	3065.000000	3065.000000	3065.000000
mean	1998.042414	6.491028	0.169759	1.146369
std	14.747378	3.456934	0.373819	1.550857
min	1973.000000	1.000000	-0.002000	0.000000
25%	1985.000000	3.000000	0.000000	0.000000
50%	1998.000000	6.000000	0.000000	0.357000
75%	2011.000000	9.000000	0.036000	1.673000
max	2024.000000	12.000000	2.047000	5.951000

	Solar Energy	Wind Energy	Wood Energy	Waste Energy
count	3065.000000	3065.000000	3065.000000	3065.000000
mean	2.015008	4.282404	36.644408	5.820124
std	5.774511	18.124793	46.900639	8.247359
min	0.000000	0.000000	0.000000	0.000000
25%	0.000000	0.000000	0.483000	0.000000
50%	0.004000	0.000000	12.062000	0.108000
75%	0.774000	0.001000	51.808000	12.764000
max	64.040000	157.409000	183.628000	32.875000

	Fuel Ethanol, Excluding Denaturant	Biomass Losses and Co-products
count	3065.000000	3065.000000
mean	6.976648	4.834706
std	21.911920	15.601717
min	0.000000	0.000000
25%	0.000000	0.000000
50%	0.007000	0.000000
75%	1.283000	0.000000
max	104.420000	75.373000

	Biomass Energy	Total Renewable Energy	Renewable Diesel Fuel
count	3065.000000	3065.000000	3065.000000
mean	46.285969	70.872209	0.428949
std	64.241520	71.197761	2.687850
min	0.000000	0.000000	0.000000

```
25%          0.258000                2.070000               0.000000
50%          9.716000               50.984000               0.000000
75%         89.359000              126.982000               0.000000
max        233.200000              308.175000              38.344000

       Other Biofuels  Conventional Hydroelectric Power    Biodiesel
count     3065.000000                       3065.000000  3065.000000
mean         0.031752                         15.757374     0.953720
std          0.258149                         32.134059     3.985003
min          0.000000                          0.000000     0.000000
25%          0.000000                          0.000000     0.000000
50%          0.000000                          0.000000     0.000000
75%          0.000000                          0.000000     0.000000
max          4.101000                        117.453000    27.871000


Global Data - Renewable Generation:
   Year  Hydro(TWh)  Biofuel(TWh)  Solar PV (TWh)  Geothermal (TWh)
0  1990     2191.67          3.88            0.09             36.42
1  1991     2268.63          4.19            0.10             37.39
2  1992     2267.16          4.63            0.12             39.30
3  1993     2397.67          5.61            0.15             40.23
4  1994     2419.73          7.31            0.17             41.05

Columns: ['Year', 'Hydro(TWh)', 'Biofuel(TWh)', 'Solar PV (TWh)', 'Geothermal
(TWh)']

Worldwide Data - Renewable Share:
   Entity Code  Year  Renewables (% equivalent primary energy)
0  Africa  NaN  1965                                  5.747495
1  Africa  NaN  1966                                  6.122062
2  Africa  NaN  1967                                  6.325731
3  Africa  NaN  1968                                  7.005293
4  Africa  NaN  1969                                  7.956088

Columns: ['Entity', 'Code', 'Year', 'Renewables (% equivalent primary energy)']

Plotting renewable generation trends…
Available columns: ['Year', 'Hydro(TWh)', 'Biofuel(TWh)', 'Solar PV (TWh)',
'Geothermal (TWh)']


Plotting renewable share evolution…
Available columns: ['Entity', 'Code', 'Year', 'Renewables (% equivalent primary
energy)']


Plotting solar and wind generation trends…
Available columns: ['Year', 'Hydro(TWh)', 'Biofuel(TWh)', 'Solar PV (TWh)',
```

```
'Geothermal (TWh)']

Available columns: ['Entity', 'Code', 'Year', 'Electricity from wind (TWh)']

Renewable Generation Data Info:

Columns: ['Year', 'Hydro(TWh)', 'Biofuel(TWh)', 'Solar PV (TWh)', 'Geothermal
(TWh)']

Sample Data:
   Year  Hydro(TWh)  Biofuel(TWh)  Solar PV (TWh)  Geothermal (TWh)
0  1990     2191.67          3.88            0.09             36.42
1  1991     2268.63          4.19            0.10             37.39
2  1992     2267.16          4.63            0.12             39.30
3  1993     2397.67          5.61            0.15             40.23
4  1994     2419.73          7.31            0.17             41.05

Latest year in data: 2017


Visualization Summary:
- Data covers years from 1990 to 2017
- Total types of renewable energy tracked: 4
- Energy types: ['Hydro(TWh)', 'Biofuel(TWh)', 'Solar PV (TWh)', 'Geothermal
(TWh)']

Starting weather impact analysis…
Weather Data Info:

Columns: ['Time', 'Energy delta[Wh]', 'GHI', 'temp', 'pressure', 'humidity',
'wind_speed', 'rain_1h', 'snow_1h', 'clouds_all', 'isSun', 'sunlightTime',
'dayLength', 'SunlightTime/daylength', 'weather_type', 'hour', 'month']

Data Types:
Time                       object
Energy delta[Wh]            int64
GHI                       float64
temp                      float64
pressure                    int64
humidity                    int64
wind_speed                float64
rain_1h                   float64
snow_1h                   float64
clouds_all                  int64
isSun                       int64
sunlightTime                int64
dayLength                   int64
SunlightTime/daylength    float64
weather_type                int64
hour                        int64
```

```
month                         int64
dtype: object


Numeric columns for analysis: ['Energy delta[Wh]', 'GHI', 'temp', 'pressure',
'humidity', 'wind_speed', 'rain_1h', 'snow_1h', 'clouds_all', 'isSun',
'sunlightTime', 'dayLength', 'SunlightTime/daylength', 'weather_type', 'hour',
'month']


Creating scatter matrix for variables: ['temp', 'wind_speed', 'GHI', 'Energy
delta[Wh]']


Summary Statistics:
                temp     wind_speed            GHI  Energy delta[Wh]
count  196776.000000  196776.000000  196776.000000     196776.000000
mean        9.790521       3.937746      32.596538        573.008228
std         7.995428       1.821694      52.172018       1044.824047
min       -16.600000       0.000000       0.000000          0.000000
25%         3.600000       2.600000       0.000000          0.000000
50%         9.300000       3.700000       1.600000          0.000000
75%        15.700000       5.000000      46.800000        577.000000
max        35.800000      14.300000     229.200000       5020.000000

Key Findings:
Correlation between temp and wind_speed: -0.08
Correlation between GHI and temp: 0.49
Correlation between GHI and wind_speed: 0.02
Correlation between Energy delta[Wh] and temp: 0.38
Correlation between Energy delta[Wh] and wind_speed: 0.03
Correlation between Energy delta[Wh] and GHI: 0.91

Starting energy mix analysis…
Renewable Generation Data Columns:
Index(['Year', 'Hydro(TWh)', 'Biofuel(TWh)', 'Solar PV (TWh)',
       'Geothermal (TWh)'],
      dtype='object')

Non-renewable Generation Data Columns:
Index(['Mode of Generation', 'Contribution (TWh)'], dtype='object')

Renewable Consumption Data Columns:
Index(['Entity', 'Code', 'Year', 'Geo Biomass Other - TWh',
       'Solar Generation - TWh', 'Wind Generation - TWh',
       'Hydro Generation - TWh'],
      dtype='object')

Total Renewable Generation: 93342.04 TWh
Total Non-renewable Generation: 38896.32 TWh
```

```
Analyzing renewable energy composition…


Renewable Energy Mix Analysis for 2017:
Hydro(TWh): 4197 TWh (71.7%)
Biofuel(TWh): 1127 TWh (19.3%)
Solar PV (TWh): 444 TWh (7.6%)
Geothermal (TWh): 85 TWh (1.5%)

Average Annual Growth Rates:
Hydro(TWh): 3.2% per year
Biofuel(TWh): 23.7% per year
Solar PV (TWh): 38.5% per year
Geothermal (TWh): 3.2% per year

Starting statistical analysis…
Renewable Generation Data Structure:

Columns: ['Year', 'Hydro(TWh)', 'Biofuel(TWh)', 'Solar PV (TWh)', 'Geothermal
(TWh)']

Sample data:
   Year  Hydro(TWh)  Biofuel(TWh)  Solar PV (TWh)  Geothermal (TWh)
0  1990     2191.67          3.88            0.09             36.42
1  1991     2268.63          4.19            0.10             37.39
2  1992     2267.16          4.63            0.12             39.30
3  1993     2397.67          5.61            0.15             40.23
4  1994     2419.73          7.31            0.17             41.05

Analyzing columns: ['Hydro(TWh)', 'Biofuel(TWh)', 'Solar PV (TWh)', 'Geothermal
(TWh)']

Growth Rates Statistics (%):
       Hydro(TWh)  Biofuel(TWh)  Solar PV (TWh)  Geothermal (TWh)
count       27.00         27.00           27.00             27.00
mean         3.21         23.69           38.47              3.23
std         13.23          8.98           21.22              2.48
min        -26.25          7.99           11.11             -2.83
25%          0.55         18.26           23.86              1.88
50%          1.62         23.08           33.33              3.15
75%          4.33         28.90           51.34              4.48
max         44.63         45.63           97.89              8.06

Variance Analysis:

                         mean         std            var          cv
Hydro(TWh)       2974.167500  595.936814  355140.686634   20.037097
Biofuel(TWh)      245.032500  329.275399  108422.288160  134.380296
```

```
Solar PV (TWh)      57.430000  113.343588   12846.768985  197.359548
Geothermal (TWh)    57.014286   14.850555     220.538996   26.047078


Summary Statistics:
Total Generation: 93342.04 TWh

Latest Year (2017) Generation Mix:
Hydro(TWh): 4197.29 TWh (71.7%)
Biofuel(TWh): 1127.31 TWh (19.3%)
Solar PV (TWh): 443.55 TWh (7.6%)
Geothermal (TWh): 85.34 TWh (1.5%)

Compound Annual Growth Rate (CAGR):
Hydro(TWh): 2.4%
Biofuel(TWh): 23.4%
Solar PV (TWh): 37.0%
Geothermal (TWh): 3.2%

<IPython.core.display.HTML object>
```