Short introduction:

Year after year, the globe becomes more modernised, and as a result, more polluted. This information was gathered from the US Energy Administration and combined for simpler examination. It's a compilation of significant components that go into CO2 Emissions, including everything from each country's production and consumption of each major energy source, as well as its annual pollution rating. Each country's GDP, population, energy intensity per capita (person), and energy intensity per GDP are also included (per person GDP). The data ranges from the 1980s through the year 2020.

Feature Descriptions:

- Country - Country in question
- Energy_type - Type of energy source
- Year - Year the data was recorded
- Energy_consumption - Amount of Consumption for the specific energy source, measured (quad Btu)
- Energy_production - Amount of Production for the specific energy source, measured (quad Btu)
- GDP - Countries GDP at purchasing power parities, measured (Billion 2015$ PPP)
- Population - Population of specific Country, measured (Mperson)
- Energyintensityper_capita - Energy intensity is a measure of the energy inefficiency of an economy. It is calculated as units of energy per unit of capita (capita = individual person), measured (MMBtu/person)
- Energyintensityby_GDP- Energy intensity is a measure of the energy inefficiency of an economy. It is calculated as units of energy per unit of GDP, measred (1000 Btu/2015$ GDP PPP)
- CO2_emission - The amount of C02 emitted, measured (MMtonnes CO2)

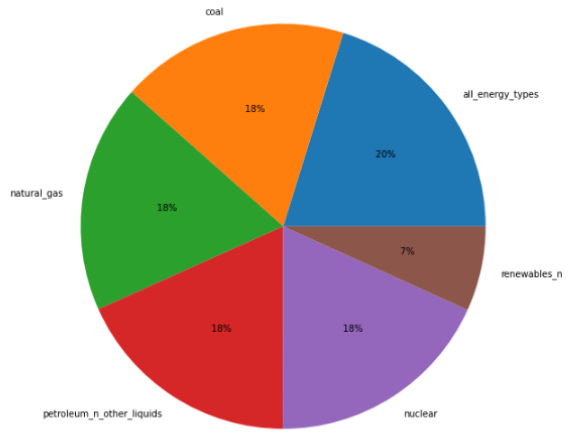
Analysis:

Step 1: First checking columns name of datasets. After having name of columns now its time to know about the information of columns data type. Information related to null values, total counts of rows.

Step 2: After knowing data type and other information its time to check the null values of dataset. There are many null values in our dataset, so we have to replace these null values and also missing values by mean, median, mode according to dataset conditions, But before we have to check the shape of dataset and its describing and correlation.

Step 3: As you can clearly see that check min, IQR, max values its better to replace vales by median, So we are replacing missing values by median of their columns. And dropping the time, longitude and latitude vales as they do not much impacting on dataset.

```
axis = plt.subplots(figsize=(10,14))
Name = df["Energy_type"].unique()
size = df["Energy_type"].value_counts()
plt.pie(size,labels=Name,autopct='%5.0f%%')
plt.show()
```



Analysis Top 10 countries according to GDP, Population, Energy production, CO2_emission. After seeing matrix china, india, USA are top 3 countries to p CO2 and Population

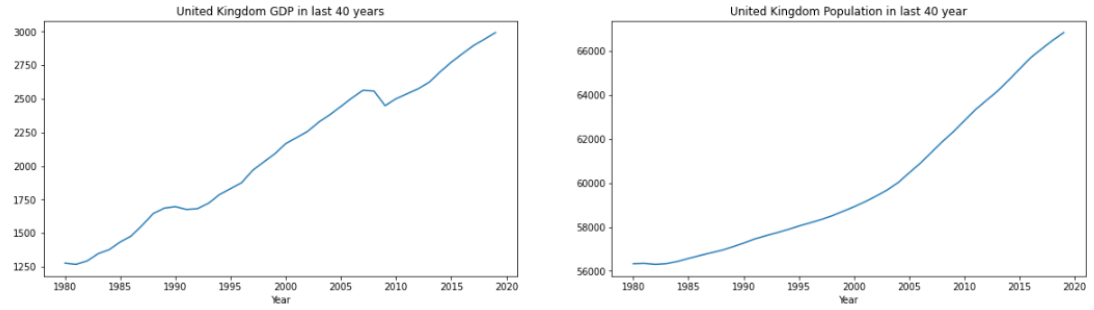
```
top_country = df[df['Year'] == max(df['Year'])].reset_index()
full_latest_top_country = top_country.groupby('Country')['GDP', 'Population', 'Energy_production', 'CO2_emissi
temp_f = full_latest_top_country.sort_values(by='Population', ascending=False)[:11]
temp_f = temp_f.reset_index(drop=True)
temp_f.style.background_gradient(cmap='Reds')
```

	Country	GDP	Population	Energy_production	CO2_emission
0	World	766141.482354	46287786.382980	1223.017937	71169.866995
1	China	138770.040000	8607072.000000	247.182719	21217.200773
2	India	55861.740000	8208840.000000	35.570744	4616.664040
3	United States	119552.580000	1980262.800000	202.801836	10554.385508
4	Indonesia	19366.002000	1625904.000000	34.118303	1058.809346
5	Pakistan	6239.280000	1299384.000000	3.855328	385.793988
6	Brazil	18247.794000	1267477.200000	25.426486	906.886544
7	Nigeria	6095.334000	1206304.800000	12.048611	202.387851
8	Bangladesh	4788.678000	978126.000000	2.218214	181.563262
9	Russia	22622.520000	875257.800000	128.555991	3596.484477
10	Mexico	14304.852000	766474.800000	11.819714	912.296677

United Kingdom GDP in last 40 years vs United Kingdom Population in last 40 year

```
# Before Adjustment
ax = plt.subplot(1, 2, 1)
dfs[dfs['Country'] == 'United Kingdom'].plot(x='Year', y='GDP', legend=None, ax=ax, figsize=(20, 5), title='United Kingdom GDP in last 40 years');

ax = plt.subplot(1, 2, 2)
dfs[dfs['Country'] == 'United Kingdom'].plot(x='Year', y='Population', legend=None, ax=ax, figsize=(20, 5), title='United Kingdom Population in last 40 year')
```



Conclusion:

1. Most CO2 emission caused by renewables and coal,natural gas and petroleum liquids. Average CO2_emission increase fastly from 1980 to 2020. Specailly last 10 years CO2 emission rapidly increases.
2. Average CO2_emission increase fastly from 1980 to 2020 by population. As population increased CO2 emission rapidly increases.
3. Analysis Top 10 countries according to GDP, Population, Energy production, CO2_emission. After seeing matrix china, india, USA are top 3 countries to produce CO2 and Population
4. Analysis Last 20 year according to GDP, Population, Energy production, CO2_emission. After seeing matrix last 3 years 2019,2018,2017 are worst year as CO2 and Population increase as compared to pervious years.

Reference:

- https://climateknowledgeportal.worldbank.org/
- https://data.worldbank.org/topic/19
- https://ieeexplore.ieee.org/document/9668116