```
print("Hi")
```

```
Hi
```

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import warnings
warnings.filterwarnings('ignore')
```

```
from google.colab import files
import matplotlib.pyplot as plt
import seaborn as sns
import pandas as pd
import numpy as np
```

```
df=pd.read_csv("/content/owid-co2-data.csv")
```

```
df.head()
```

	iso_code	country	year	co2	consumption_co2	co2_growth_prct	co2_growth_abs	trade_co2	co2_per_capita	consumption_co2_pe
0	AFG	Afghanistan	1949	0.015	NaN	NaN	NaN	NaN	0.002	
1	AFG	Afghanistan	1950	0.084	NaN	475.0	0.070	NaN	0.011	
2	AFG	Afghanistan	1951	0.092	NaN	8.7	0.007	NaN	0.012	
3	AFG	Afghanistan	1952	0.092	NaN	0.0	0.000	NaN	0.012	
4	AFG	Afghanistan	1953	0.106	NaN	16.0	0.015	NaN	0.013	

5 rows × 58 columns

```
print("shape",df.shape)
```

```
shape (25204, 58)
```

```
print("columns:",df.columns.tolist()[:58])
```

```
columns: ['iso_code', 'country', 'year', 'co2', 'consumption_co2', 'co2_growth_prct', 'co2_growth_abs', 'trade_co2', 'co2_per_capita
```

```
display(df.info())
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 25204 entries, 0 to 25203
Data columns (total 58 columns):
 #   Column                                Non-Null Count  Dtype
---  ------                                --------------  -----
 0   iso_code                              21948 non-null  object
 1   country                               25204 non-null  object
 2   year                                  25204 non-null  int64
 3   co2                                   23949 non-null  float64
 4   consumption_co2                       3976 non-null   float64
 5   co2_growth_prct                       24931 non-null  float64
 6   co2_growth_abs                        23585 non-null  float64
 7   trade_co2                             3976 non-null   float64
 8   co2_per_capita                        23307 non-null  float64
 9   consumption_co2_per_capita            3976 non-null   float64
 10  share_global_co2                      23949 non-null  float64
 11  cumulative_co2                        23949 non-null  float64
 12  share_global_cumulative_co2           23949 non-null  float64
 13  co2_per_gdp                           15389 non-null  float64
 14  consumption_co2_per_gdp               3761 non-null   float64
 15  co2_per_unit_energy                   9141 non-null   float64
 16  coal_co2                              17188 non-null  float64
 17  cement_co2                            12248 non-null  float64
 18  flaring_co2                           4382 non-null   float64
 19  gas_co2                               8845 non-null   float64
 20  oil_co2                               20539 non-null  float64
 21  other_industry_co2                    1999 non-null   float64
 22  cement_co2_per_capita                 12218 non-null  float64
 23  coal_co2_per_capita                   16860 non-null  float64
 24  flaring_co2_per_capita                4381 non-null   float64
 25  gas_co2_per_capita                    8835 non-null   float64
 26  oil_co2_per_capita                    20181 non-null  float64
 27  other_co2_per_capita                  1999 non-null   float64
 28  trade_co2_share                       3976 non-null   float64
 29  share_global_cement_co2               12248 non-null  float64
 30  share_global_coal_co2                 17188 non-null  float64
 31  share_global_flaring_co2              4382 non-null   float64
 32  share_global_gas_co2                  8845 non-null   float64
 33  share_global_oil_co2                  20539 non-null  float64
 34  share_global_other_co2                1999 non-null   float64
 35  cumulative_cement_co2                 12248 non-null  float64
 36  cumulative_coal_co2                   17188 non-null  float64
 37  cumulative_flaring_co2                4382 non-null   float64
 38  cumulative_gas_co2                    8845 non-null   float64
 39  cumulative_oil_co2                    20539 non-null  float64
 40  cumulative_other_co2                  1999 non-null   float64
 41  share_global_cumulative_cement_co2    12248 non-null  float64
 42  share_global_cumulative_coal_co2      17188 non-null  float64
 43  share_global_cumulative_flaring_co2   4382 non-null   float64
 44  share_global_cumulative_gas_co2       8845 non-null   float64
 45  share_global_cumulative_oil_co2       20539 non-null  float64
 46  share_global_cumulative_other_co2     1999 non-null   float64
 47  total_ghg                             5208 non-null   float64
 48  ghg_per_capita                        5155 non-null   float64
 49  methane                               5211 non-null   float64
 50  methane_per_capita                    5157 non-null   float64
 51  nitrous_oxide                         5211 non-null   float64
 52  nitrous_oxide_per_capita              5157 non-null   float64
 53  population                            22878 non-null  float64
 54  gdp                                   13538 non-null  float64
 55  primary_energy_consumption            8690 non-null   float64
 56  energy_per_capita                     8681 non-null   float64
 57  energy_per_gdp                        6803 non-null   float64
dtypes: float64(55), int64(1), object(2)
memory usage: 11.2+ MB
None
```

```
#Missing values summary
missing_values = df.isnull().sum().sort_values(ascending=False)
print(missing_values)
```

```
other_industry_co2                    23205
other_co2_per_capita                  23205
share_global_cumulative_other_co2     23205
cumulative_other_co2                  23205
share_global_other_co2                23205
consumption_co2_per_gdp               21443
consumption_co2_per_capita            21228
consumption_co2                       21228
trade_co2                             21228
trade_co2_share                       21228
flaring_co2_per_capita                20823
cumulative_flaring_co2                20822
share_global_flaring_co2              20822
share_global_cumulative_flaring_co2   20822
flaring_co2                           20822
ghg_per_capita                        20049
methane_per_capita                    20047
```

```
nitrous_oxide_per_capita               20047
total_ghg                              19996
nitrous_oxide                          19993
methane                                19993
energy_per_gdp                         18401
energy_per_capita                      16523
primary_energy_consumption             16514
gas_co2_per_capita                     16369
gas_co2                                16359
share_global_gas_co2                   16359
cumulative_gas_co2                     16359
share_global_cumulative_gas_co2        16359
co2_per_unit_energy                    16063
cement_co2_per_capita                  12986
cumulative_cement_co2                  12956
share_global_cumulative_cement_co2     12956
share_global_cement_co2                12956
cement_co2                             12956
gdp                                    11666
co2_per_gdp                             9815
coal_co2_per_capita                     8344
share_global_coal_co2                   8016
share_global_cumulative_coal_co2        8016
cumulative_coal_co2                     8016
coal_co2                                8016
oil_co2_per_capita                      5023
cumulative_oil_co2                      4665
oil_co2                                 4665
share_global_oil_co2                    4665
share_global_cumulative_oil_co2         4665
iso_code                                3256
population                              2326
co2_per_capita                          1897
co2_growth_abs                          1619
co2                                     1255
cumulative_co2                          1255
share_global_cumulative_co2             1255
share_global_co2                        1255
co2_growth_prct                          273
country                                    0
```

```
missing_percentage=(missing_values/len(df)*100).round(2)
print(missing_percentage)
```

```
other_co2_per_capita                   92.07
share_global_cumulative_other_co2      92.07
cumulative_other_co2                   92.07
share_global_other_co2                 92.07
consumption_co2_per_gdp                85.08
consumption_co2_per_capita             84.22
consumption_co2                        84.22
trade_co2                              84.22
trade_co2_share                        84.22
flaring_co2_per_capita                 82.62
cumulative_flaring_co2                 82.61
share_global_flaring_co2               82.61
share_global_cumulative_flaring_co2    82.61
flaring_co2                            82.61
ghg_per_capita                         79.55
methane_per_capita                     79.54
nitrous_oxide_per_capita               79.54
total_ghg                              79.34
nitrous_oxide                          79.32
methane                                79.32
energy_per_gdp                         73.01
energy_per_capita                      65.56
primary_energy_consumption             65.52
gas_co2_per_capita                     64.95
gas_co2                                64.91
share_global_gas_co2                   64.91
cumulative_gas_co2                     64.91
share_global_cumulative_gas_co2        64.91
co2_per_unit_energy                    63.73
cement_co2_per_capita                  51.52
cumulative_cement_co2                  51.40
share_global_cumulative_cement_co2     51.40
share_global_cement_co2                51.40
cement_co2                             51.40
gdp                                    46.29
co2_per_gdp                            38.94
coal_co2_per_capita                    33.11
share_global_coal_co2                  31.80
share_global_cumulative_coal_co2       31.80
cumulative_coal_co2                    31.80
coal_co2                               31.80
oil_co2_per_capita                     19.93
cumulative_oil_co2                     18.51
oil_co2                                18.51
```

```
share_global_cumulative_oil_co2        18.51
iso_code                               12.92
population                              9.23
co2_per_capita                          7.53
co2_growth_abs                          6.42
co2                                     4.98
cumulative_co2                          4.98
share_global_cumulative_co2             4.98
share_global_co2                        4.98
co2_growth_prct                         1.08
country                                 0.00
year                                    0.00
dtype: float64
```

```
missing_df=pd.DataFrame({'missing_values':missing_values,'missing_percentage':missing_percentage})
print(missing_df)
```

```
other_industry_co2                          23205              92.07
other_co2_per_capita                        23205              92.07
share_global_cumulative_other_co2           23205              92.07
cumulative_other_co2                        23205              92.07
share_global_other_co2                      23205              92.07
consumption_co2_per_gdp                     21443              85.08
consumption_co2_per_capita                  21228              84.22
consumption_co2                             21228              84.22
trade_co2                                   21228              84.22
trade_co2_share                             21228              84.22
flaring_co2_per_capita                      20823              82.62
cumulative_flaring_co2                      20822              82.61
share_global_flaring_co2                    20822              82.61
share_global_cumulative_flaring_co2         20822              82.61
flaring_co2                                 20822              82.61
ghg_per_capita                              20049              79.55
methane_per_capita                          20047              79.54
nitrous_oxide_per_capita                    20047              79.54
total_ghg                                   19996              79.34
nitrous_oxide                               19993              79.32
methane                                     19993              79.32
energy_per_gdp                              18401              73.01
energy_per_capita                           16523              65.56
primary_energy_consumption                  16514              65.52
gas_co2_per_capita                          16369              64.95
gas_co2                                     16359              64.91
share_global_gas_co2                        16359              64.91
cumulative_gas_co2                          16359              64.91
share_global_cumulative_gas_co2             16359              64.91
co2_per_unit_energy                         16063              63.73
cement_co2_per_capita                       12986              51.52
cumulative_cement_co2                       12956              51.40
share_global_cumulative_cement_co2          12956              51.40
share_global_cement_co2                     12956              51.40
cement_co2                                  12956              51.40
gdp                                         11666              46.29
co2_per_gdp                                  9815              38.94
coal_co2_per_capita                          8344              33.11
share_global_coal_co2                        8016              31.80
share_global_cumulative_coal_co2             8016              31.80
cumulative_coal_co2                          8016              31.80
coal_co2                                     8016              31.80
oil_co2_per_capita                           5023              19.93
cumulative_oil_co2                           4665              18.51
oil_co2                                      4665              18.51
share_global_oil_co2                         4665              18.51
share_global_cumulative_oil_co2              4665              18.51
iso_code                                     3256              12.92
population                                   2326               9.23
co2_per_capita                               1897               7.53
co2_growth_abs                               1619               6.42
co2                                          1255               4.98
cumulative_co2                               1255               4.98
share_global_cumulative_co2                  1255               4.98
share_global_co2                             1255               4.98
co2_growth_prct                               273               1.08
country                                         0               0.00
year                                            0               0.00
```

```
print(df.describe())
```

```
std       53.791369   1521.680894      3373.348275       702.565991
min     1750.000000      0.000000         0.197000       -99.640000
25%     1925.000000      0.528000        10.319750        -0.450000
50%     1967.000000      4.857000        57.091000         3.350000
75%     1995.000000     42.818000       276.378750        10.460000
max     2020.000000  36702.503000     36702.503000    102318.510000

       co2_growth_abs    trade_co2  co2_per_capita  \
count    23585.000000  3976.000000    23307.000000
```

```
25%         -0.611000    -0.892500        0.253000
50%          0.059000     1.953000        1.250000
75%          1.103000     9.700500        4.657500
max       1736.258000  1028.487000      748.639000

       consumption_co2_per_capita  share_global_co2  cumulative_co2  ...  \
count                 3976.000000      23949.000000    2.394900e+04  ...
mean                     6.567721          4.983979    1.035710e+04  ...
std                      6.927957         17.704994    6.120603e+04  ...
min                      0.055000          0.000000    0.000000e+00  ...
25%                      1.240250          0.010000    6.992000e+00  ...
50%                      4.359500          0.060000    9.131800e+01  ...
75%                      9.848500          0.600000    1.147514e+03  ...
max                     57.792000        100.000000    1.696524e+06  ...

       ghg_per_capita      methane  methane_per_capita  nitrous_oxide  \
count     5155.000000  5211.000000         5157.000000    5211.000000
mean         8.099506    82.253370            2.046923      29.308135
std          9.387477   566.702756            3.534126     199.291118
min        -50.487000     0.000000            0.000000       0.000000
25%          2.514000     2.150000            0.715000       0.535000
50%          5.474000     9.030000            1.111000       3.590000
75%         10.299000    31.055000            1.690000      10.570000
max         86.991000  8660.010000           39.812000    3054.000000

       nitrous_oxide_per_capita    population           gdp  \
count               5157.000000  2.287800e+04  1.353800e+04
mean                   0.606690  7.072322e+07  2.877088e+11
std                    0.844671  3.795858e+08  2.180094e+12
min                    0.000000  1.490000e+03  5.543200e+07
25%                    0.228000  1.291899e+06  9.828867e+09
50%                    0.384000  4.880320e+06  3.037019e+10
75%                    0.607000  1.759622e+07  1.268944e+11
max                    8.239000  7.794799e+09  1.136302e+14

       primary_energy_consumption  energy_per_capita  energy_per_gdp
count                 8690.000000        8681.000000     6803.000000
mean                  1569.083895       25568.487454        1.850291
std                   9066.102799       33319.956208        1.575161
min                      0.000000           0.000000        0.050000
25%                      6.997000        3270.369000        0.856000
50%                     61.404500       13701.318000        1.407000
75%                    352.884500       35493.775000        2.351000
max                 162194.290000      317582.498000       13.493000

[8 rows x 56 columns]
```

```
#Check number of unique countries and year range
print("Unique Countries :", df['country'].unique())
```

```
Unique Countries : ['Afghanistan' 'Africa' 'Albania' 'Algeria' 'Andorra' 'Angola' 'Anguilla'
 'Antarctica' 'Antigua and Barbuda' 'Argentina' 'Armenia' 'Aruba' 'Asia'
 'Asia (excl. China & India)' 'Australia' 'Austria' 'Azerbaijan' 'Bahamas'
 'Bahrain' 'Bangladesh' 'Barbados' 'Belarus' 'Belgium' 'Belize' 'Benin'
 'Bermuda' 'Bhutan' 'Bolivia' 'Bonaire Sint Eustatius and Saba'
 'Bosnia and Herzegovina' 'Botswana' 'Brazil' 'British Virgin Islands'
 'Brunei' 'Bulgaria' 'Burkina Faso' 'Burundi' 'Cambodia' 'Cameroon'
 'Canada' 'Cape Verde' 'Central African Republic' 'Chad' 'Chile' 'China'
 'Christmas Island' 'Colombia' 'Comoros' 'Congo' 'Cook Islands'
 'Costa Rica' "Cote d'Ivoire" 'Croatia' 'Cuba' 'Curacao' 'Cyprus'
 'Czechia' 'Democratic Republic of Congo' 'Denmark' 'Djibouti' 'Dominica'
 'Dominican Republic' 'EU-27' 'EU-28' 'Ecuador' 'Egypt' 'El Salvador'
 'Equatorial Guinea' 'Eritrea' 'Estonia' 'Eswatini' 'Ethiopia' 'Europe'
 'Europe (excl. EU-27)' 'Europe (excl. EU-28)' 'Faeroe Islands' 'Fiji'
 'Finland' 'France' 'French Equatorial Africa' 'French Guiana'
 'French Polynesia' 'French West Africa' 'Gabon' 'Gambia' 'Georgia'
 'Germany' 'Ghana' 'Greece' 'Greenland' 'Grenada' 'Guadeloupe' 'Guatemala'
 'Guinea' 'Guinea-Bissau' 'Guyana' 'Haiti' 'Honduras' 'Hong Kong'
 'Hungary' 'Iceland' 'India' 'Indonesia' 'International transport' 'Iran'
 'Iraq' 'Ireland' 'Israel' 'Italy' 'Jamaica' 'Japan' 'Jordan' 'Kazakhstan'
 'Kenya' 'Kiribati' 'Kosovo' 'Kuwait' 'Kuwaiti Oil Fires' 'Kyrgyzstan'
 'Laos' 'Latvia' 'Lebanon' 'Leeward Islands' 'Lesotho' 'Liberia' 'Libya'
 'Liechtenstein' 'Lithuania' 'Luxembourg' 'Macao' 'Madagascar' 'Malawi'
 'Malaysia' 'Maldives' 'Mali' 'Malta' 'Marshall Islands' 'Martinique'
 'Mauritania' 'Mauritius' 'Mayotte' 'Mexico' 'Micronesia'
 'Micronesia (country)' 'Moldova' 'Mongolia' 'Montenegro' 'Montserrat'
 'Morocco' 'Mozambique' 'Myanmar' 'Namibia' 'Nauru' 'Nepal' 'Netherlands'
 'New Caledonia' 'New Zealand' 'Nicaragua' 'Niger' 'Nigeria' 'Niue'
 'North America' 'North America (excl. USA)' 'North Korea'
 'North Macedonia' 'Norway' 'Oceania' 'Oman' 'Pakistan' 'Palau'
 'Palestine' 'Panama' 'Panama Canal Zone' 'Papua New Guinea' 'Paraguay'
 'Peru' 'Philippines' 'Poland' 'Portugal' 'Puerto Rico' 'Qatar' 'Reunion'
 'Romania' 'Russia' 'Rwanda' 'Ryukyu Islands' 'Saint Helena'
 'Saint Kitts and Nevis' 'Saint Lucia' 'Saint Pierre and Miquelon'
 'Saint Vincent and the Grenadines' 'Samoa' 'Sao Tome and Principe'
 'Saudi Arabia' 'Senegal' 'Serbia' 'Seychelles' 'Sierra Leone' 'Singapore'
 'Sint Maarten (Dutch part)' 'Slovakia' 'Slovenia' 'Solomon Islands'
 'Somalia' 'South Africa' 'South America' 'South Korea' 'South Sudan'
 'Spain' 'Sri Lanka' 'St. Kitts-Nevis-Anguilla' 'Sudan' 'Suriname'
```

```
       'Sweden' 'Switzerland' 'Syria' 'Taiwan' 'Tajikistan' 'Tanzania'
       'Thailand' 'Timor' 'Togo' 'Tonga' 'Trinidad and Tobago' 'Tunisia'
       'Turkey' 'Turkmenistan' 'Turks and Caicos Islands' 'Tuvalu' 'Uganda'
       'Ukraine' 'United Arab Emirates' 'United Kingdom' 'United States'
       'Uruguay' 'Uzbekistan' 'Vanuatu' 'Venezuela' 'Vietnam'
       'Wallis and Futuna' 'World' 'Yemen' 'Zambia' 'Zimbabwe']
```

```
print("Year range:",df['year'].min(),"to",df['year'].max())
```

```
Year range: 1750 to 2020
```

```
before_dup=df.shape[0]
df=df.drop_duplicates()
print("Dropped_duplicates: ",before_dup - df.shape[0])
```

```
Dropped_duplicates:  0
```

```
keep_cols = ['iso_code','country','year','co2','co2_per_capita','gdp','population','coal_co2','oil_co2','gas_co2']
keep_cols=[c for c in keep_cols if c in df.columns]
df=df[keep_cols].copy()
print("Columns kept :",df.columns.tolist())
```

```
Columns kept : ['iso_code', 'country', 'year', 'co2', 'co2_per_capita', 'gdp', 'population', 'coal_co2', 'oil_co2', 'gas_co2']
```

```
before_drop = df.shape[0]
df = df.dropna(subset=['country','year','co2'])
print("Dropped rows missing country/year/co2:", before_drop - df.shape[0])
print("Remaining rows:", df.shape[0])
```

```
Dropped rows missing country/year/co2: 1255
Remaining rows: 23949
```

```
if 'co2_per_capita' in df.columns and 'population' in df.columns:
    mask = (
        df['co2_per_capita'].isnull()
        & df['co2'].notnull()
        & df['population'].notnull()
        & (df['population'] > 0)
    )
    filled = mask.sum()
    df.loc[mask, 'co2_per_capita'] = df.loc[mask, 'co2'] / df.loc[mask, 'population']
    print(f"Filled co2_per_capita for {filled} rows using co2/population")
```

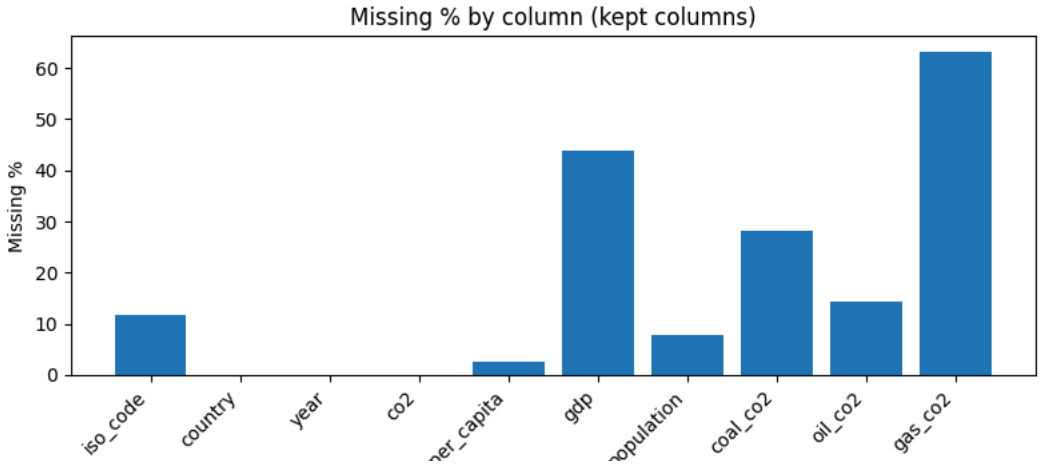
```
Filled co2_per_capita for 49 rows using co2/population
```

```
#convert the year into the int datatype
df['year']=df['year'].astype(int)
```

```
miss_pct=(df.isnull().sum()/len(df)*100).round(2)
print("Missing %  by columns (kept cols):")
print(miss_pct)
```

```
Missing %  by columns (kept cols):
iso_code          11.85
country            0.00
year               0.00
co2                0.00
co2_per_capita     2.48
gdp               43.94
population         7.72
coal_co2          28.23
oil_co2           14.24
gas_co2           63.07
dtype: float64
```

```
plt.figure(figsize=(8,4))
plt.bar(miss_pct.index.astype(str), miss_pct.values)
plt.xticks(rotation=45, ha='right')
plt.ylabel('Missing %')
plt.title('Missing % by column (kept columns)')
plt.tight_layout()
plt.show()
```


Missing % by column (kept columns)
Missing %
0
10
20
30
40
50
60
iso_code
country
year
co2
gdp
coal_co2
oil_co2
gas_co2