In [1]:
```python
import pandas as pd
dataset=pd.read_csv('world-data-2023.csv')
dataset
```

Out[1]:

	Country	Density\n(P/Km2)	Abbreviation	Agricultural Land(%)	Land Area(Km2)	Armed Forces size	Birth Rate	Call C
0	Afghanistan	60	AF	58.10%	652,230	323,000	32.49	9
1	Albania	105	AL	43.10%	28,748	9,000	11.78	35
2	Algeria	18	DZ	17.40%	2,381,741	317,000	24.28	21
3	Andorra	164	AD	40.00%	468	NaN	7.20	37
4	Angola	26	AO	47.50%	1,246,700	117,000	40.73	24
...	...	...	...	...	...	...	...	
190	Venezuela	32	VE	24.50%	912,050	343,000	17.88	5
191	Vietnam	314	VN	39.30%	331,210	522,000	16.75	8
192	Yemen	56	YE	44.60%	527,968	40,000	30.45	96
193	Zambia	25	ZM	32.10%	752,618	16,000	36.19	26
194	Zimbabwe	38	ZW	41.90%	390,757	51,000	30.68	26

195 rows × 35 columns

In [2]:
```python
type(dataset)
```

Out[2]: pandas.core.frame.DataFrame

In [3]:
```
dataset.info
```

```
Out[3]: <bound method DataFrame.info of          Country Density\n(P/Km2) Abbrevia
        tion Agricultural Land( %)  \
        0    Afghanistan              60           AF              58.10%
        1        Albania             105           AL              43.10%
        2        Algeria              18           DZ              17.40%
        3        Andorra             164           AD              40.00%
        4         Angola              26           AO              47.50%
        ..           ...             ...          ...                 ...
        190    Venezuela              32           VE              24.50%
        191      Vietnam             314           VN              39.30%
        192        Yemen              56           YE              44.60%
        193       Zambia              25           ZM              32.10%
        194     Zimbabwe              38           ZW              41.90%

            Land Area(Km2) Armed Forces size  Birth Rate  Calling Code  \
        0          652,230           323,000       32.49          93.0
        1           28,748             9,000       11.78         355.0
        2        2,381,741           317,000       24.28         213.0
        3              468               NaN        7.20         376.0
        4        1,246,700           117,000       40.73         244.0
        ..             ...               ...         ...           ...
        190        912,050           343,000       17.88          58.0
        191        331,210           522,000       16.75          84.0
        192        527,968            40,000       30.45         967.0
        193        752,618            16,000       36.19         260.0
        194        390,757            51,000       30.68         263.0

            Capital/Major City Co2-Emissions  ... Out of pocket health expenditure
        \
        0                Kabul         8,672  ...                           78.40%
        1               Tirana         4,536  ...                           56.90%
        2              Algiers       150,006  ...                           28.10%
        3     Andorra la Vella           469  ...                           36.40%
        4               Luanda        34,693  ...                           33.40%
        ..                 ...           ...  ...                              ...
        190            Caracas       164,175  ...                           45.80%
        191              Hanoi       192,668  ...                           43.50%
        192              Sanaa        10,609  ...                           81.00%
        193             Lusaka         5,141  ...                           27.50%
        194             Harare        10,983  ...                           25.80%

            Physicians per thousand  Population  \
        0                      0.28  38,041,754
        1                      1.20   2,854,191
        2                      1.72  43,053,054
        3                      3.33      77,142
        4                      0.21  31,825,295
        ..                      ...         ...
        190                    1.92  28,515,829
        191                    0.82  96,462,106
        192                    0.31  29,161,922
        193                    1.19  17,861,030
        194                    0.21  14,645,468

             Population: Labor force participation (%) Tax revenue (%) Total tax r
        ate  \
        0                                       48.90%            9.30%        71.
        40%
        1                                       55.70%           18.60%        36.
        60%
        2                                       41.20%           37.20%        66.
```

```
10%
3                                            NaN             NaN
NaN
4                                         77.50%           9.20%          49.
10%
..                                           ...             ...
...
190                                       59.70%             NaN          73.
30%
191                                       77.40%          19.10%          37.
60%
192                                       38.00%             NaN          26.
60%
193                                       74.60%          16.20%          15.
60%
194                                       83.10%          20.70%          31.
60%

    Unemployment rate Urban_population   Latitude   Longitude
0              11.12%        9,797,273  33.939110   67.709953
1              12.33%        1,747,593  41.153332   20.168331
2              11.70%       31,510,100  28.033886    1.659626
3                 NaN           67,873  42.506285    1.521801
4               6.89%       21,061,025 -11.202692   17.873887
..                ...              ...        ...         ...
190             8.80%       25,162,368   6.423750  -66.589730
191             2.01%       35,332,140  14.058324  108.277199
192            12.91%       10,869,523  15.552727   48.516388
193            11.43%        7,871,713 -13.133897   27.849332
194             4.95%        4,717,305 -19.015438   29.154857

[195 rows x 35 columns]>
```

In [4]:
```
dataset.describe()
```

Out[4]:

	Birth Rate	Calling Code	Fertility Rate	Infant mortality	Life expectancy	Maternal mortality ratio	Physicians per thousand
count	189.000000	194.000000	188.000000	189.000000	187.000000	181.000000	188.000000
mean	20.214974	360.546392	2.698138	21.332804	72.279679	160.392265	1.839840
std	9.945774	323.236419	1.282267	19.548058	7.483661	233.502024	1.684261
min	5.900000	1.000000	0.980000	1.400000	52.800000	2.000000	0.010000
25%	11.300000	82.500000	1.705000	6.000000	67.000000	13.000000	0.332500
50%	17.950000	255.500000	2.245000	14.000000	73.200000	53.000000	1.460000
75%	28.750000	506.750000	3.597500	32.700000	77.500000	186.000000	2.935000
max	46.080000	1876.000000	6.910000	84.500000	85.400000	1150.000000	8.420000

In [5]:
```
dataset=dataset.drop_duplicates()
dataset
```
Out[5]:

	Country	Density\n(P/Km2)	Abbreviation	Agricultural Land(%)	Land Area(Km2)	Armed Forces size	Birth Rate	Call C
0	Afghanistan	60	AF	58.10%	652,230	323,000	32.49	9
1	Albania	105	AL	43.10%	28,748	9,000	11.78	35
2	Algeria	18	DZ	17.40%	2,381,741	317,000	24.28	21
3	Andorra	164	AD	40.00%	468	NaN	7.20	37
4	Angola	26	AO	47.50%	1,246,700	117,000	40.73	24
...	...	...	...	...	...	...	...	
190	Venezuela	32	VE	24.50%	912,050	343,000	17.88	5
191	Vietnam	314	VN	39.30%	331,210	522,000	16.75	8
192	Yemen	56	YE	44.60%	527,968	40,000	30.45	96
193	Zambia	25	ZM	32.10%	752,618	16,000	36.19	26
194	Zimbabwe	38	ZW	41.90%	390,757	51,000	30.68	26

195 rows × 35 columns

In [6]:
```
dataset.isnull()
```
Out[6]:

	Country	Density\n(P/Km2)	Abbreviation	Agricultural Land(%)	Land Area(Km2)	Armed Forces size	Birth Rate	Calling Code
0	False	False	False	False	False	False	False	False
1	False	False	False	False	False	False	False	False
2	False	False	False	False	False	False	False	False
3	False	False	False	False	False	True	False	False
4	False	False	False	False	False	False	False	False
...	...	...	...	...	...	...	...	...
190	False	False	False	False	False	False	False	False
191	False	False	False	False	False	False	False	False
192	False	False	False	False	False	False	False	False
193	False	False	False	False	False	False	False	False
194	False	False	False	False	False	False	False	False

195 rows × 35 columns

In [7]:
```
dataset.isnull().sum()
```

Out[7]:
```
Country                                       0
Density\n(P/Km2)                              0
Abbreviation                                  7
Agricultural Land( %)                         7
Land Area(Km2)                                1
Armed Forces size                            24
Birth Rate                                    6
Calling Code                                  1
Capital/Major City                            3
Co2-Emissions                                 7
CPI                                          17
CPI Change (%)                               16
Currency-Code                                15
Fertility Rate                                7
Forested Area (%)                             7
Gasoline Price                               20
GDP                                           2
Gross primary education enrollment (%)        7
Gross tertiary education enrollment (%)      12
Infant mortality                              6
Largest city                                  6
Life expectancy                               8
Maternal mortality ratio                     14
Minimum wage                                 45
Official language                             5
Out of pocket health expenditure              7
Physicians per thousand                       7
Population                                    1
Population: Labor force participation (%)    19
Tax revenue (%)                              26
Total tax rate                               12
Unemployment rate                            19
Urban_population                              5
Latitude                                      1
Longitude                                     1
dtype: int64
```

In [8]:
```
dataset.notnull()
```

Out[8]:

	Country	Density\n(P/Km2)	Abbreviation	Agricultural Land(%)	Land Area(Km2)	Armed Forces size	Birth Rate	Calling Code
0	True	True	True	True	True	True	True	True
1	True	True	True	True	True	True	True	True
2	True	True	True	True	True	True	True	True
3	True	True	True	True	True	False	True	True
4	True	True	True	True	True	True	True	True
...	...	...	...	...	...	...	...	...
190	True	True	True	True	True	True	True	True
191	True	True	True	True	True	True	True	True
192	True	True	True	True	True	True	True	True
193	True	True	True	True	True	True	True	True
194	True	True	True	True	True	True	True	True

195 rows × 35 columns

In [9]:
```
dataset.isnull().sum().sum()
```

Out[9]: 341

In [11]:
```
df=dataset.fillna(value=0)
df
```

Out[11]:

	Country	Density\n(P/Km2)	Abbreviation	Agricultural Land(%)	Land Area(Km2)	Armed Forces size	Birth Rate	Call C
0	Afghanistan	60	AF	58.10%	652,230	323,000	32.49	9
1	Albania	105	AL	43.10%	28,748	9,000	11.78	35
2	Algeria	18	DZ	17.40%	2,381,741	317,000	24.28	21
3	Andorra	164	AD	40.00%	468	0	7.20	37
4	Angola	26	AO	47.50%	1,246,700	117,000	40.73	24
...	...	...	...	...	...	...	...	
190	Venezuela	32	VE	24.50%	912,050	343,000	17.88	5
191	Vietnam	314	VN	39.30%	331,210	522,000	16.75	8
192	Yemen	56	YE	44.60%	527,968	40,000	30.45	96
193	Zambia	25	ZM	32.10%	752,618	16,000	36.19	26
194	Zimbabwe	38	ZW	41.90%	390,757	51,000	30.68	26

195 rows × 35 columns

In [12]:
```
df1=dataset.fillna(method='pad')
df1
```

Out[12]:

	Country	Density\n(P/Km2)	Abbreviation	Agricultural Land(%)	Land Area(Km2)	Armed Forces size	Birth Rate	Call C
0	Afghanistan	60	AF	58.10%	652,230	323,000	32.49	9
1	Albania	105	AL	43.10%	28,748	9,000	11.78	35
2	Algeria	18	DZ	17.40%	2,381,741	317,000	24.28	21
3	Andorra	164	AD	40.00%	468	317,000	7.20	37
4	Angola	26	AO	47.50%	1,246,700	117,000	40.73	24
...	...	...	...	...	...	...	...	
190	Venezuela	32	VE	24.50%	912,050	343,000	17.88	5
191	Vietnam	314	VN	39.30%	331,210	522,000	16.75	8
192	Yemen	56	YE	44.60%	527,968	40,000	30.45	96
193	Zambia	25	ZM	32.10%	752,618	16,000	36.19	26
194	Zimbabwe	38	ZW	41.90%	390,757	51,000	30.68	26

195 rows × 35 columns

In [13]:
```python
df2=dataset.fillna(method='bfill')
df2
```

Out[13]:

	Country	Density\n(P/Km2)	Abbreviation	Agricultural Land(%)	Land Area(Km2)	Armed Forces size	Birth Rate	Call C
0	Afghanistan	60	AF	58.10%	652,230	323,000	32.49	9
1	Albania	105	AL	43.10%	28,748	9,000	11.78	35
2	Algeria	18	DZ	17.40%	2,381,741	317,000	24.28	21
3	Andorra	164	AD	40.00%	468	117,000	7.20	37
4	Angola	26	AO	47.50%	1,246,700	117,000	40.73	24
...	...	...	...	...	...	...	...	
190	Venezuela	32	VE	24.50%	912,050	343,000	17.88	5
191	Vietnam	314	VN	39.30%	331,210	522,000	16.75	8
192	Yemen	56	YE	44.60%	527,968	40,000	30.45	96
193	Zambia	25	ZM	32.10%	752,618	16,000	36.19	26
194	Zimbabwe	38	ZW	41.90%	390,757	51,000	30.68	26

195 rows × 35 columns

In [14]:
```python
import numpy as np
import matplotlib.pyplot as plt
from scipy import stats
```

In [15]:
```python
#detecting the outliers using IQR
df.columns
```

Out[15]:
```
Index(['Country', 'Density\n(P/Km2)', 'Abbreviation', 'Agricultural Land(
%)',
       'Land Area(Km2)', 'Armed Forces size', 'Birth Rate', 'Calling Cod
e',
       'Capital/Major City', 'Co2-Emissions', 'CPI', 'CPI Change (%)',
       'Currency-Code', 'Fertility Rate', 'Forested Area (%)',
       'Gasoline Price', 'GDP', 'Gross primary education enrollment (%)',
       'Gross tertiary education enrollment (%)', 'Infant mortality',
       'Largest city', 'Life expectancy', 'Maternal mortality ratio',
       'Minimum wage', 'Official language', 'Out of pocket health expendit
ure',
       'Physicians per thousand', 'Population',
       'Population: Labor force participation (%)', 'Tax revenue (%)',
       'Total tax rate', 'Unemployment rate', 'Urban_population', 'Latitud
e',
       'Longitude'],
      dtype='object')
```

In [16]:
```
df.drop(['Country', 'Density\n(P/Km2)', 'Abbreviation', 'Agricultural Land(
df.columns
```

Out[16]:
```
Index(['Birth Rate', 'Calling Code', 'Fertility Rate', 'Infant mortality',
       'Life expectancy', 'Maternal mortality ratio',
       'Physicians per thousand', 'Latitude', 'Longitude'],
      dtype='object')
```

In [17]:
```
Q1=df.quantile(0.25)
Q3=df.quantile(0.75)
IQR=Q3-Q1
print(IQR)
```

```
Birth Rate                   17.770000
Calling Code                425.000000
Fertility Rate                1.940000
Infant mortality             26.550000
Life expectancy              11.100000
Maternal mortality ratio    166.000000
Physicians per thousand       2.630000
Latitude                     35.733221
Longitude                    55.705194
dtype: float64
```

In [18]:
```
df=df[~((df<(Q1-1.5*IQR))|(df>(Q3+1.5*IQR))).any(axis=1)]
df
```

Out[18]:

	Birth Rate	Calling Code	Fertility Rate	Infant mortality	Life expectancy	Maternal mortality ratio	Physicians per thousand	Latitude	Longit
1	11.78	355.0	1.62	7.8	78.5	15.0	1.20	41.153332	20.168
2	24.28	213.0	3.02	20.1	76.7	112.0	1.72	28.033886	1.659
4	40.73	244.0	5.52	51.6	60.8	241.0	0.21	-11.202692	17.873
5	15.33	1.0	1.99	5.0	76.9	42.0	2.76	17.060816	-61.796
6	17.02	54.0	2.26	8.8	76.5	39.0	3.96	-38.416097	-63.616
...	...	...	...	...	...	...	...	...	
188	23.30	998.0	2.42	19.1	71.6	29.0	2.37	41.377491	64.585
190	17.88	58.0	2.27	21.4	72.1	125.0	1.92	6.423750	-66.589
191	16.75	84.0	2.05	16.5	75.3	43.0	0.82	14.058324	108.277
192	30.45	967.0	3.79	42.9	66.1	164.0	0.31	15.552727	48.516
193	36.19	260.0	4.63	40.4	63.5	213.0	1.19	-13.133897	27.849

145 rows × 9 columns

In [19]:
```
df.describe()
```

Out[19]:

	Birth Rate	Calling Code	Fertility Rate	Infant mortality	Life expectancy	Maternal mortality ratio	Physicians per thousand
count	145.000000	145.000000	145.000000	145.000000	145.000000	145.000000	145.000000
mean	18.076345	352.834483	2.376759	16.589655	74.098621	86.648276	2.029310
std	8.454403	323.776415	1.021711	14.724647	6.005738	107.260362	1.528937
min	6.400000	1.000000	0.980000	0.000000	58.400000	0.000000	0.000000
25%	10.600000	60.000000	1.620000	5.000000	70.900000	9.000000	0.710000
50%	16.750000	256.000000	2.060000	12.200000	74.900000	37.000000	1.920000
75%	22.460000	504.000000	2.790000	24.400000	78.100000	129.000000	3.070000
max	40.730000	998.000000	5.520000	62.600000	85.400000	401.000000	6.350000

In [ ]: