

import all required libraries for data analysis

```
In [1]: import pandas as pd
import seaborn as sns
import numpy as np
import matplotlib.pyplot as plt
```

import data file

```
In [4]: data = pd.read_csv('solar_power_by_country.csv')
```

		2016_New	2016_Total	2017_New	2017_Total	2018_New	2018_Total	2019_New	2019_Total	2020_New	2020_Total	W per capita 2019	Share of total consumption %
0	China	34540.0	78070.0	53000.0	131000.0	45000.0	175018	30100.0	204700	49655.0	254355	147.0	6.2
1	European Union	0.0	101433.0	0.0	107150.0	8300.0	115234	16000.0	134129	18788.0	152917	295.0	6.0
2	United States	14730.0	40300.0	10600.0	51000.0	10600.0	53184	13300.0	60682	14890.0	75572	231.0	3.4
3	Japan	8600.0	42750.0	7000.0	49000.0	6500.0	55500	7000.0	63000	4000.0	67000	498.0	8.3
4	Germany	1520.0	41220.0	1800.0	42000.0	3000.0	45930	3900.0	49200	4583.0	53783	593.0	9.7
...
75	Oman	0.0	2.0	0.0	8.0	0.0	8	0.0	9	0.0	109	0.0	0.0
76	Colombia	0.0	2.0	0.0	11.0	0.0	86	0.0	90	0.0	107	0.0	0.0
77	Kenya	0.0	32.0	0.0	39.0	0.0	105	0.0	106	0.0	106	0.0	0.0
78	Guatemala	0.0	93.0	0.0	99.0	0.0	101	0.0	101	0.0	101	0.0	0.0
79	Croatia	8.0	56.0	4.0	60.0	1.0	61	0.0	69	0.0	85	17.0	0.0

80 rows × 13 columns

head() function

```
In [6]: data.head()
```

		2016_New	2016_Total	2017_New	2017_Total	2018_New	2018_Total	2019_New	2019_Total	2020_New	2020_Total	W per capita 2019	Share of total consumption %
0	China	34540.0	78070.0	53000.0	131000.0	45000.0	175018	30100.0	204700	49655.0	254355	147.0	6.2
1	European Union	NaN	101433.0	NaN	107150.0	8300.0	115234	16000.0	134129	18788.0	152917	295.0	6.0
2	United States	14730.0	40300.0	10600.0	51000.0	10600.0	53184	13300.0	60682	14890.0	75572	231.0	3.4
3	Japan	8600.0	42750.0	7000.0	49000.0	6500.0	55500	7000.0	63000	4000.0	67000	498.0	8.3
4	Germany	1520.0	41220.0	1800.0	42000.0	3000.0	45930	3900.0	49200	4583.0	53783	593.0	9.7

tail() function

```
In [7]: data.tail()
```

		2016_New	2016_Total	2017_New	2017_Total	2018_New	2018_Total	2019_New	2019_Total	2020_New	2020_Total	W per capita 2019	Share of total consumption %
75	Oman	NaN	2.0	NaN	8.0	NaN	8	NaN	9	NaN	109	NaN	NaN
76	Colombia	NaN	2.0	NaN	11.0	NaN	86	NaN	90	NaN	107	NaN	NaN
77	Kenya	NaN	32.0	NaN	39.0	NaN	105	NaN	106	NaN	106	NaN	NaN
78	Guatemala	NaN	93.0	NaN	99.0	NaN	101	NaN	101	NaN	101	NaN	NaN
79	Croatia	8.0	56.0	4.0	60.0	1.0	61	NaN	69	NaN	85	17.0	NaN

```
In [9]: data.tail(10)
```

		2016_New	2016_Total	2017_New	2017_Total	2018_New	2018_Total	2019_New	2019_Total	2020_New	2020_Total	W per capita 2019	Share of total consumption %
70	Lithuania	1.0	76.0	4.0	74.0	10.0	84	NaN	103	NaN	148	37.0	NaN
71	Namibia	NaN	30.0	NaN	70.0	NaN	88	NaN	135	NaN	145	55.0	NaN
72	New Zealand	NaN	53.0	NaN	70.0	NaN	90	NaN	117	NaN	142	NaN	NaN
73	Estonia	NaN	10.0	NaN	15.0	NaN	32	NaN	121	NaN	130	NaN	NaN
74	Bolivia	NaN	6.0	NaN	8.0	NaN	70	NaN	120	NaN	120	NaN	NaN
75	Oman	NaN	2.0	NaN	8.0	NaN	8	NaN	9	NaN	109	NaN	NaN
76	Colombia	NaN	2.0	NaN	11.0	NaN	86	NaN	90	NaN	107	NaN	NaN
77	Kenya	NaN	32.0	NaN	39.0	NaN	105	NaN	106	NaN	106	NaN	NaN
78	Guatemala	NaN	93.0	NaN	99.0	NaN	101	NaN	101	NaN	101	NaN	NaN
79	Croatia	8.0	56.0	4.0	60.0	1.0	61	NaN	69	NaN	85	17.0	NaN

```
In [10]: data.shape
```

(80, 13)

```
In [11]: data.describe()
```

	2016_New	2016_Total	2017_New	2017_Total	2018_New	2018_Total	2019_New	2019_Total	2020_New	2020_Total	W per capita 2019	Share of total consumption %
count	36.000000	75.000000	36.000000	75.000000	20.000000	80.000000	14.000000	80.000000	27.000000	80.000000	59.000000	42.000000
mean	2038.527778	5284.040000	2596.138889	6608.546667	4882.255000	7486.962500	6839.428571	8900.562500	5228.333333	10757.000000	141.101695	4.690476
std	6235.740879	16523.151061	8988.374584	21201.180765	10076.187123	24952.807349	8312.978405	28970.430989	9948.967340	34808.938642	145.559408	3.169736
min	1.000000	2.000000	4.000000	8.000000	0.000000	8.000000	81.000000	9.000000	15.000000	85.000000	2.000000	0.100000
25%	52.500000	49.500000	59.250000	99.500000	89.775000	134.000000	1209.500000	210.750000	736.500000	214.000000	34.500000	2.500000
50%	185.000000	202.000000	211.500000	298.000000	1450.000000	490.500000	3800.000000	855.000000	1833.000000	1060.500000	88.000000	3.900000
75%	748.500000	1530.000000	900.000000	1850.000000	4475.000000	2573.750000	9175.000000	3372.500000	4061.000000	5431.000000	190.000000	6.500000
max	34540.00000	101433.00000	53000.00000	131000.00000	45000.00000	175018.00000	30100.00000	204700.00000	49655.00000	254355.00000	637.000000	12.900000

all the data set columns are here

```
In [23]: data.columns
```

```
Out[23]: Index(['Country or territory', '2016_New', '2016_Total', '2017_New',
      '2017_Total', '2018_New', '2018_Total', '2019_New', '2019_Total',
      '2020_New', '2020_Total', 'W per capita 2019', 'Share of total consumption %'],
      dtype='object')
```

```
In [24]: data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 80 entries, 0 to 79
Data columns (total 13 columns):
 #   Column                Non-Null Count  Dtype
---  --
 0   Country or territory    80 non-null     object
 1   2016_New                80 non-null     float64
 2   2016_Total              80 non-null     float64
 3   2017_New                80 non-null     float64
 4   2017_Total              80 non-null     float64
 5   2018_New                80 non-null     float64
 6   2018_Total              80 non-null     int64
 7   2019_New                80 non-null     float64
 8   2019_Total              80 non-null     int64
 9   2020_New                80 non-null     float64
10   2020_Total              80 non-null     int64
11   W per capita 2019       80 non-null     float64
12   Share of total consumption % 80 non-null     float64
dtypes: float64(9), int64(3), object(1)
memory usage: 8.2+ KB
```

finding null values in the data set

```
In [12]: data.isnull()
```

		2016_New	2016_Total	2017_New	2017_Total	2018_New	2018_Total	2019_New	2019_Total	2020_New	2020_Total	W per capita 2019	Share of total consumption %
0		False	False	False	False	False	False	False	False	False	False	False	False
1		False	True	False	True	False	False	False	False	False	False	False	False
2		False	False	False	False	False	False	False	False	False	False	False	False
3		False	False	False	False	False	False	False	False	False	False	False	False
4		False	False	False	False	False	False	False	False	False	False	False	False
...
75		False	True	False	True	False	True	False	True	False	False	True	True
76		False	True	False	True	False	True	False	True	False	True	False	True
77		False	True	False	True	False	True	False	True	False	True	False	True
78		False	True	False	True	False	True	False	True	False	True	False	True
79		False	False	False	False	False	False	True	False	True	False	False	True

80 rows × 13 columns

```
In [13]: data.isnull().sum()
```

Country or territory	0
2016_New	44
2016_Total	5
2017_New	44
2017_Total	5
2018_New	60
2018_Total	0
2019_New	66
2019_Total	0
2020_New	0
2020_Total	53
W per capita 2019	21
Share of total consumption %	38
dtype: int64	

data cleaning process are as below type

```
In [16]: data.fillna(0, inplace=True)
data.isnull().sum()
```

Country or territory	0
2016_New	0
2016_Total	0
2017_New	0
2017_Total	0
2018_New	0
2018_Total	0
2019_New	0
2019_Total	0
2020_New	0
2020_Total	0
W per capita 2019	0
Share of total consumption %	0
dtype: int64	

all the columns data type

```
In [17]: data.dtypes
```

Country or territory	object
2016_New	float64
2016_Total	float64
2017_New	float64
2017_Total	float64
2018_New	float64
2018_Total	int64
2019_New	float64
2019_Total	int64
2020_New	float64
2020_Total	int64
W per capita 2019	float64
Share of total consumption %	float64
dtype: object	

```
In [18]: data
```

		2016_New	2016_Total	2017_New	2017_Total	2018_New	2018_Total	2019_New	2019_Total	2020_New	2020_Total	W per capita 2019	Share of total consumption %
0	China	34540.0	78070.0	53000.0	131000.0	45000.0	175018	30100.0	204700	49655.0	254355	147.0	6.2
1	European Union	0.0	101433.0	0.0	107150.0	8300.0	115234	16000.0	134129	18788.0	152917	295.0	6.0
2	United States	14730.0	40300.0	10600.0	51000.0	10600.0	53184	13300.0	60682	14890.0	75572	231.0	3.4
3	Japan	8600.0	42750.0	7000.0	49000.0	6500.0	55500	7000.0	63000	4000.0	67000	498.0	8.3
4	Germany	1520.0	41220.0	1800.0	42000.0	3000.0	45930	3900.0	49200	4583.0	53783	593.0	9.7
...
75	Oman	0.0	2.0	0.0	8.0	0.0	8	0.0	9	0.0	109	0.0	0.0
76	Colombia	0.0	2.0	0.0	11.0	0.0	86	0.0	90	0.0	107	0.0	0.0
77	Kenya	0.0	32.0	0.0	39.0	0.0	105	0.0	106	0.0	106	0.0	0.0
78	Guatemala	0.0	93.0	0.0	99.0	0.0	101	0.0	101	0.0	101	0.0	0.0
79	Croatia	8.0	56.0	4.0	60.0	1.0	61	0.0	69	0.0	85	17.0	0.0

80 rows × 13 columns

country has highest total value in 2020

```
In [20]: data[data['2020_Total'] == data['2020_Total'].max()]
```

		2016_New	2016_Total	2017_New	2017_Total	2018_New	2018_Total	2019_New	2019_Total	2020_New	2020_Total	W per capita 2019	Share of total consumption %
0	China	34540.0	78070.0	53000.0	131000.0	45000.0	175018	30100.0	204700	49655.0	254355	147.0	6.2

country has highest total value in 2017

```
In [21]: data[data['2017_Total'] == data['2017_Total'].max()]
```

		2016_New	2016_Total	2017_New	2017_Total	2018_New	2018_Total	2019_New	2019_Total	2020_New	2020_Total	W per capita 2019	Share of total consumption %
0	China	34540.0	78070.0	53000.0	131000.0	45000.0	175018	30100.0	204700	49655.0	254355	147.0	6.2

Countries with a score higher than 5000 in 2019

```
In [22]: data[data['2019_Total']>5000]
```

0	China	34540.0	78070.0	53000.0	131000.0	45000.0	175018	30100.0	204700	49655.0	254355	147.0	6.2
1	European Union	0.0	101433.0	0.0	107150.0	8300.0	115234	16000.0	134129	18788.0	152917	295.0	6.0
2	United States	14730.0	40300.0	10600.0	51000.0	10600.0	53184	13300.0	60682	14890.0	75572	231.0	3.4
3	Japan	8600.0	42750.0	7000.0	49000.0	6500.0	55500	7000.0	63000	4000.0	67000	498.0	8.3
4	Germany	1520.0	41220.0	1800.0	42000.0	3000.0	45930	3900.0	49200	4583.0	53783	593.0	9.7
5	India	3970.0	9010.0	9100.0	18300.0	10800.0	26869	9900.0	35089	4122.0	39211	32.0	6.5
6	Italy	373.0	19279.0	409.0	19700.0	0.0	20120	600.0	20080	800.0	21600	345.0	8.3
7	Australia	839.0	5900.0	1250.0	7200.0	3800.0	11300	3700.0	15928	1699.0	17627	637.0	10.7
8	Vietnam	0.0	6.0	0.0	9.0	0.0	106	4800.0	5695	10909.0	16504	60.0	0.0
9	South Korea	850.0	4350.0	1200.0	5600.0	2000.0	7862	3100.0	11200	3375.0	14575	217.0	3.8
10	Spain	0.0	4669.0	0.0	4688.0	0.0	4707	0.0	8711	5378.0	14089	186.0	9.0
11	United Kingdom	1970.0	11630.0	900.0	12700.0	0.0	13108	233.0	13346	177.0	13563	200.0	4.0
12	France	559.0	7130.0	875.0	8000.0	0.0	9483	900.0	9900	1833.0	11733	148.0	2.8
13	Netherlands	525.0	2100.0	853.0	2900.0	1300.0	4150	0.0	6725	3488.0	10213	396.0	8.9
15	Turkey	584.0	832.0	2600.0	3400.0	1600.0	5063	0.0	5995	673.0	6668	73.0	5.9

this is boxplot graph

```
data.boxplot()  
  
#<AxesSubplot>
```