PANDAS CASE STUDY

WE WILL CHECK DATA FAO STAT

In [7]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
df = pd.read_csv(r"C:\Users\Asad\Desktop\Data_Assignment\FAOSTAT_data_1-12-202
```

In [8]:
```
# Basis statistics or summary of data
df.describe()
```

Out[8]:

	Area Code	Element Code	Item Code	Year Code	Year	Value
count	180.0	180.000000	180.0	180.000000	180.000000	1.800000e+02
mean	231.0	5413.666667	1717.0	1990.500000	1990.500000	1.235340e+08
std	0.0	81.146808	0.0	17.366409	17.366409	1.427124e+08
min	231.0	5312.000000	1717.0	1961.000000	1961.000000	2.522300e+04
25%	231.0	5312.000000	1717.0	1975.750000	1975.750000	6.438125e+04
50%	231.0	5419.000000	1717.0	1990.500000	1990.500000	6.076251e+07
75%	231.0	5510.000000	1717.0	2005.250000	2005.250000	2.376391e+08
max	231.0	5510.000000	1717.0	2020.000000	2020.000000	5.034661e+08

In [9]:
```
df.head()
```

Out[9]:

	Domain Code	Domain	Area Code	Area	Element Code	Element	Item Code	Item	Year Code	Year	Unit	V
0	QCL	Crops and livestock products	231	United States of America	5312	Area harvested	1717	Cereals, Total	1961	1961	ha	6486
1	QCL	Crops and livestock products	231	United States of America	5419	Yield	1717	Cereals, Total	1961	1961	hg/ha	2
2	QCL	Crops and livestock products	231	United States of America	5510	Production	1717	Cereals, Total	1961	1961	tonnes	16361

	Domain Code	Domain	Area Code	Area	Element Code	Element	Item Code	Item	Year Code	Year	Unit	V
3	QCL	Crops and livestock products	231	United States of America	5312	Area harvested	1717	Cereals, Total	1962	1962	ha	6054

In [10]:

```
# Install csv viewer , or excel viewer an we can check file by using csv viewe

# how can we do basic commands on this FAO Dataset
df.head(5)
```

Out[10]:

	Domain Code	Domain	Area Code	Area	Element Code	Element	Item Code	Item	Year Code	Year	Unit	V
0	QCL	Crops and livestock products	231	United States of America	5312	Area harvested	1717	Cereals, Total	1961	1961	ha	6486
1	QCL	Crops and livestock products	231	United States of America	5419	Yield	1717	Cereals, Total	1961	1961	hg/ha	2
2	QCL	Crops and livestock products	231	United States of America	5510	Production	1717	Cereals, Total	1961	1961	tonnes	16361
3	QCL	Crops and livestock products	231	United States of America	5312	Area harvested	1717	Cereals, Total	1962	1962	ha	6054
4	QCL	Crops and livestock products	231	United States of America	5419	Yield	1717	Cereals, Total	1962	1962	hg/ha	2

In [13]:

```
new_df=df.drop(['Flag' , 'Flag Description' , 'Item' , ] , axis=1)
new_df.head(5)
```

Out[13]:

	Domain Code	Domain	Area Code	Area	Element Code	Element	Item Code	Year Code	Year	Unit	Value
0	QCL	Crops and livestock products	231	United States of America	5312	Area harvested	1717	1961	1961	ha	64869558

	Domain Code	Domain	Area Code	Area	Element Code	Element	Item Code	Year Code	Year	Unit	Value
1	QCL	Crops and livestock products	231	United States of America	5419	Yield	1717	1961	1961	hg/ha	25223
2	QCL	Crops and livestock products	231	United States of America	5510	Production	1717	1961	1961	tonnes	163619978
3	QCL	Crops and livestock products	231	United States of America	5312	Area harvested	1717	1962	1962	ha	60548008
	Crops	United									

In [14]:
```
# writing in the end of code has similiar meaning instead of saving in new na
df.drop(['Flag' , 'Flag Description' , 'Item' , ] , axis=1).head()
```

Out[14]:

	Domain Code	Domain	Area Code	Area	Element Code	Element	Item Code	Year Code	Year	Unit	Value
0	QCL	Crops and livestock products	231	United States of America	5312	Area harvested	1717	1961	1961	ha	64869558
1	QCL	Crops and livestock products	231	United States of America	5419	Yield	1717	1961	1961	hg/ha	25223
2	QCL	Crops and livestock products	231	United States of America	5510	Production	1717	1961	1961	tonnes	163619978
3	QCL	Crops and livestock products	231	United States of America	5312	Area harvested	1717	1962	1962	ha	60548008
4	QCL	Crops and livestock products	231	United States of America	5419	Yield	1717	1962	1962	hg/ha	26831

In [16]:
```
df.mean()
```

Out[16]:
```
Area Code        2.310000e+02
Element Code     5.413667e+03
Item Code        1.717000e+03
Year Code        1.990500e+03
Year             1.990500e+03
Value            1.235340e+08
```

dtype: float64

In [23]:
```
df.groupby(["Element", 'Unit']).mean()
```

Out[23]:

Element	Unit	Area Code	Element Code	Item Code	Year Code	Year	Value
Area harvested	**ha**	231.0	5312.0	1717.0	1990.5	1990.5	6.200303e+07
Production	**tonnes**	231.0	5510.0	1717.0	1990.5	1990.5	3.085482e+08
Yield	**hg/ha**	231.0	5419.0	1717.0	1990.5	1990.5	5.063017e+04

In [26]:
```
df.value_counts(["Element"])
```

Out[26]:
```
Element
Area harvested    60
Production        60
Yield             60
dtype: int64
```

In [27]:
```
df.groupby(["Element", 'Unit']).mean()
```

Out[27]:

Element	Unit	Area Code	Element Code	Item Code	Year Code	Year	Value
Area harvested	**ha**	231.0	5312.0	1717.0	1990.5	1990.5	6.200303e+07
Production	**tonnes**	231.0	5510.0	1717.0	1990.5	1990.5	3.085482e+08
Yield	**hg/ha**	231.0	5419.0	1717.0	1990.5	1990.5	5.063017e+04

In [40]:
```
# use of bolen operator

df[df['Value'] >= 10000].groupby(["Element", 'Unit']).mean()
```

Out[40]:

Element	Unit	Area Code	Element Code	Item Code	Year Code	Year	Value
Area harvested	**ha**	231.0	5312.0	1717.0	1990.5	1990.5	6.200303e+07
Production	**tonnes**	231.0	5510.0	1717.0	1990.5	1990.5	3.085482e+08
Yield	**hg/ha**	231.0	5419.0	1717.0	1990.5	1990.5	5.063017e+04

In [ ]:
```

```