In [4]:

```
import pandas as pd
import numpy as np
```

In [5]:

```
ataFrame(np.arange(0,20).reshape(5,4),index=['Row1','Row2','Row3','Row4','Row5','Row6'],columns=["Column1","Column2","Column3","Column4"])
```

In [15]:

```
df.tail()
```

Out[15]:

	Column1	Column2	Column3	Coumn4
Row1	0	1	2	3
Row2	4	5	6	7
Row3	8	9	10	11
Row4	12	13	14	15
Row5	16	17	18	19

In [7]:

```
df.to_csv('sample.csv')
```

In [8]:

```
df.loc['Row1']
```

Out[8]:

```
Column1    0
Column2    1
Column3    2
Coumn4     3
Name: Row1, dtype: int32
```

In [9]:

```
df.loc[['Row1','Row2']]
```

Out[9]:

	Column1	Column2	Column3	Coumn4
Row1	0	1	2	3
Row2	4	5	6	7

In [ ]:

```
type(['df.locRow1'])
```

In [23]:

```
df=pd.read_csv('C:/Users/DGVC/Downloads/crop1.csv')
```

In [24]:

```
df.head()
```

Out[24]:

	Area	Item	Element	Year	Unit	Value
0	Afghanistan	Almonds, with shell	Area harvested	1975	ha	0.0
1	Afghanistan	Almonds, with shell	Area harvested	1976	ha	5900.0
2	Afghanistan	Almonds, with shell	Area harvested	1977	ha	6000.0
3	Afghanistan	Almonds, with shell	Area harvested	1978	ha	6000.0
4	Afghanistan	Almonds, with shell	Area harvested	1979	ha	6000.0

In [25]:

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1895975 entries, 0 to 1895974
Data columns (total 6 columns):
 #   Column   Dtype  
---  ------   -----  
 0   Area     object 
 1   Item     object 
 2   Element  object 
 3   Year     int64  
 4   Unit     object 
 5   Value    float64
dtypes: float64(1), int64(1), object(4)
memory usage: 86.8+ MB
```

In [26]:

```
df.describe()
```

Out[26]:

	Year	Value
count	1.895975e+06	1.766475e+06
mean	1.992321e+03	1.077275e+06
std	1.718118e+01	1.403138e+07
min	1.961000e+03	0.000000e+00
25%	1.978000e+03	4.500000e+03
50%	1.994000e+03	2.687500e+04
75%	2.007000e+03	1.303910e+05
max	2.020000e+03	1.955308e+09

In [28]:

```
df['Item'].value_counts()
```

Out[28]:

```
Vegetables, fresh nes    38635
Maize                    33379
Tomatoes                 33326
Potatoes                 31597
Fruit, fresh nes         31485
                         ...  
Cinnamon (cannella)       4238
Cloves                    4070
Tung nuts                 3338
Ramie                     2943
Tallowtree seed            774
Name: Item, Length: 118, dtype: int64
```

In [30]:

```
df[df['Year']>100]
```

Out[30]:

	Area	Item	Element	Year	Unit	Value
0	Afghanistan	Almonds, with shell	Area harvested	1975	ha	0.0
1	Afghanistan	Almonds, with shell	Area harvested	1976	ha	5900.0
2	Afghanistan	Almonds, with shell	Area harvested	1977	ha	6000.0
3	Afghanistan	Almonds, with shell	Area harvested	1978	ha	6000.0
4	Afghanistan	Almonds, with shell	Area harvested	1979	ha	6000.0
...	...	...	...	...	...	...
1895970	Net Food Importing Developing Countries	Wheat	Production	2016	tonnes	52845182.0
1895971	Net Food Importing Developing Countries	Wheat	Production	2017	tonnes	57099662.0
1895972	Net Food Importing Developing Countries	Wheat	Production	2018	tonnes	55571780.0
1895973	Net Food Importing Developing Countries	Wheat	Production	2019	tonnes	53903035.0
1895974	Net Food Importing Developing Countries	Wheat	Production	2020	tonnes	54159916.0

1895975 rows × 6 columns

In [31]:

```
df.corr()
```

Out[31]:

	Year	Value
Year	1.000000	0.014543
Value	0.014543	1.000000

In [32]:

```
lst_data=[[1,2,3],[3,4,np.nan],[5,6,np.nan],[np.nan,np.nan,np.nan]]
```

In [33]:

```
type(lst_data)
```

Out[33]:

```
list
```

In [34]:

```
df=pd.DataFrame(lst_data)
```

In [35]:

```
df.head()
```

Out[35]:

	0	1	2
0	1.0	2.0	3.0
1	3.0	4.0	NaN
2	5.0	6.0	NaN
3	NaN	NaN	NaN

In [36]:

```
df.dropna(axis=0)
```

Out[36]:

	0	1	2
0	1.0	2.0	3.0

In [37]:

```
df.dropna(axis=1)
```

Out[37]:

0
1
2
3

In [38]:

```
df = pd.DataFrame(np.random.randn(5, 3), index=['a', 'c', 'e', 'f', 'h'],
                  columns=['one', 'two', 'three'])
```

In [39]:

```
df.head()
```

Out[39]:

	one	two	three
a	0.661659	0.261772	-0.808783
c	-0.308071	0.078853	-0.135434
e	-0.061840	-0.855599	-0.932110
f	0.072725	1.584597	0.268969
h	1.447613	0.374623	0.646683

In [40]:

```
df2=df.reindex(['a', 'b', 'c', 'd', 'e', 'f', 'g', 'h'])
```

In [41]:

```
df2
```

Out[41]:

	one	two	three
a	0.661659	0.261772	-0.808783
b	NaN	NaN	NaN
c	-0.308071	0.078853	-0.135434
d	NaN	NaN	NaN
e	-0.061840	-0.855599	-0.932110
f	0.072725	1.584597	0.268969
g	NaN	NaN	NaN
h	1.447613	0.374623	0.646683

In [42]:

```
df2.dropna(axis=0)
```

Out[42]:

	one	two	three
a	0.661659	0.261772	-0.808783
c	-0.308071	0.078853	-0.135434
e	-0.061840	-0.855599	-0.932110
f	0.072725	1.584597	0.268969
h	1.447613	0.374623	0.646683

In [43]:

```
pd.isna(df2['one'])
```

Out[43]:

```
a    False
b     True
c    False
d     True
e    False
f    False
g     True
h    False
Name: one, dtype: bool
```

In [44]:

```
df2['one'].notna()
```

Out[44]:

```
a     True
b    False
c     True
d    False
e     True
f     True
g    False
h     True
Name: one, dtype: bool
```

In [45]:

```
df2.fillna('Missing')
```

Out[45]:

	one	two	three
a	0.661659	0.261772	-0.808783
b	Missing	Missing	Missing
c	-0.308071	0.0788531	-0.135434
d	Missing	Missing	Missing
e	-0.0618395	-0.855599	-0.93211
f	0.0727245	1.5846	0.268969
g	Missing	Missing	Missing
h	1.44761	0.374623	0.646683

In [46]:

```
df2['one'].values
```

Out[46]:

```
array([ 0.66165929,         nan, -0.3080707 ,         nan, -0.06183952,
        0.07272454,         nan,  1.44761273])
```

In [47]:

```
df2['one'].unique
```

Out[47]:

```
<bound method Series.unique of a    0.661659
b         NaN
c   -0.308071
d         NaN
e   -0.061840
f    0.072725
g         NaN
h    1.447613
Name: one, dtype: float64>
```

In [48]:

```
df2['one'].shape
```

Out[48]:

```
(8,)
```

In [49]:

```
df.mean()
```

Out[49]:

```
one      0.362417
two      0.288849
three   -0.192135
dtype: float64
```

In [50]:

```
df.median()
```

Out[50]:

```
one      0.072725
two      0.261772
three   -0.135434
dtype: float64
```

In [51]:

```
df.count()
```

Out[51]:

```
one      5
two      5
three    5
dtype: int64
```

In [52]:

```
df.min()
```

Out[52]:

```
one      -0.308071
two      -0.855599
three    -0.932110
dtype: float64
```

In [53]:

```
df.max()
```

Out[53]:

```
one       1.447613
two       1.584597
three     0.646683
dtype: float64
```

In [55]:

```
df.corr()
```

Out[55]:

	one	two	three
one	1.000000	0.136888	0.453836
two	0.136888	1.000000	0.641777
three	0.453836	0.641777	1.000000

In [56]:

```
df.isnull()
```

Out[56]:

	one	two	three
a	False	False	False
c	False	False	False
e	False	False	False
f	False	False	False
h	False	False	False

In [57]:

```
df.notnull()
```

Out[57]:

	one	two	three
a	True	True	True
c	True	True	True
e	True	True	True
f	True	True	True
h	True	True	True

In [58]:

```
df.tail()
```

Out[58]:

	one	two	three
a	0.661659	0.261772	-0.808783
c	-0.308071	0.078853	-0.135434
e	-0.061840	-0.855599	-0.932110
f	0.072725	1.584597	0.268969
h	1.447613	0.374623	0.646683

In [ ]:

```

```