In [1]:
```
#DataFrame Groupby Examples
import pandas as pd
```

In [2]:
```
df=pd.read_csv("D:\\Python-Workspace\\Pandas\\peoples.csv")
print(df,type(df))
```

```
   CID      CNAME   CITY  EXP
0  100      SUSMA     TS   10
1  200   SAGATIKA     AP   15
2  300      LAXMI     TS   10
3  400     ROSSUM     AP   20
4  500     NARESH  DELHI   10
5  600      VINAY  DELHI   15
6  700  TEJASWANI     AP   15
7  800     FAZINA     TS   15
8  900    SANSKAR  DELHI   20
9  750        KVR     AP   10 <class 'pandas.core.frame.DataFrame'>
```

In [3]:
```
#get the City Based Group information
```

In [4]:
```
grp=df.groupby("CITY")
print("Type of group=",type(grp))
```

```
Type of group= <class 'pandas.core.groupby.generic.DataFrameGroupBy'>
```

In [5]:
```
for grpname,grpinfo in grp:
    print("-"*50)
    print("Group Name=",grpname)
    print("Records")
    print(grpinfo)
    print("-"*50)
```

```
--------------------------------------------------
Group Name= AP
Records
   CID      CNAME CITY  EXP
1  200   SAGATIKA   AP   15
3  400     ROSSUM   AP   20
6  700  TEJASWANI   AP   15
9  750        KVR   AP   10
--------------------------------------------------
--------------------------------------------------
Group Name= DELHI
Records
   CID    CNAME   CITY  EXP
4  500   NARESH  DELHI   10
5  600    VINAY  DELHI   15
8  900  SANSKAR  DELHI   20
--------------------------------------------------
--------------------------------------------------
Group Name= TS
Records
   CID   CNAME CITY  EXP
0  100   SUSMA   TS   10
2  300   LAXMI   TS   10
7  800  FAZINA   TS   15
--------------------------------------------------
```

In [6]:
```
for grpname,grpinfo in df.groupby("CITY"):
    print("-"*50)
    print("Group Name=",grpname)
    print("Records")
    print(grpinfo)
    print("-"*50)
```

```
--------------------------------------------------
Group Name= AP
Records
   CID      CNAME CITY  EXP
1  200   SAGATIKA   AP   15
3  400     ROSSUM   AP   20
6  700  TEJASWANI   AP   15
9  750        KVR   AP   10
--------------------------------------------------
--------------------------------------------------
Group Name= DELHI
Records
   CID    CNAME   CITY  EXP
4  500   NARESH  DELHI   10
5  600    VINAY  DELHI   15
8  900  SANSKAR  DELHI   20
--------------------------------------------------
--------------------------------------------------
Group Name= TS
Records
   CID   CNAME CITY  EXP
0  100   SUSMA   TS   10
2  300   LAXMI   TS   10
7  800  FAZINA   TS   15
--------------------------------------------------
```

In [7]:
```
for grpname,grpinfo in df.groupby("EXP"):
    print("-"*50)
    print("Group Name=",grpname)
    print("Records")
    print(grpinfo)
    print("-"*50)
```

```
--------------------------------------------------
Group Name= 10
Records
   CID   CNAME   CITY  EXP
0  100   SUSMA     TS   10
2  300   LAXMI     TS   10
4  500  NARESH  DELHI   10
9  750     KVR     AP   10
--------------------------------------------------
--------------------------------------------------
Group Name= 15
Records
   CID      CNAME   CITY  EXP
1  200   SAGATIKA     AP   15
5  600      VINAY  DELHI   15
6  700  TEJASWANI     AP   15
7  800     FAZINA     TS   15
--------------------------------------------------
--------------------------------------------------
Group Name= 20
Records
   CID    CNAME   CITY  EXP
3  400   ROSSUM     AP   20
8  900  SANSKAR  DELHI   20
--------------------------------------------------
```

In [8]:
```
grp=df.groupby("CITY")
grp.count()
```

Out[8]:

CITY	CID	CNAME	EXP
AP	4	4	4
DELHI	3	3	3
TS	3	3	3

In [9]:
```
df.groupby("EXP").count()
```

Out[9]:

EXP	CID	CNAME	CITY
10	4	4	4
15	4	4	4
20	2	2	2

In [10]:
```
for grpname,grpinfo in df.groupby("CITY"):
    print("-"*50)
    print("Group Name=",grpname)
    print("Records")
    print(grpinfo)
    print("-"*50)
```

```
--------------------------------------------------
Group Name= AP
Records
   CID      CNAME CITY  EXP
1  200   SAGATIKA   AP   15
3  400     ROSSUM   AP   20
6  700  TEJASWANI   AP   15
9  750        KVR   AP   10
--------------------------------------------------
--------------------------------------------------
Group Name= DELHI
Records
   CID    CNAME   CITY  EXP
4  500   NARESH  DELHI   10
5  600    VINAY  DELHI   15
8  900  SANSKAR  DELHI   20
--------------------------------------------------
--------------------------------------------------
Group Name= TS
Records
   CID   CNAME CITY  EXP
0  100   SUSMA   TS   10
2  300   LAXMI   TS   10
7  800  FAZINA   TS   15
--------------------------------------------------
```

In [11]:
```
grp=df.groupby("CITY")
grp.first()
```

Out[11]:

CITY	CID	CNAME	EXP
AP	200	SAGATIKA	15
DELHI	500	NARESH	10
TS	100	SUSMA	10

In [12]:
```
grp=df.groupby("CITY")
grp.last()
```

Out[12]:

CITY	CID	CNAME	EXP
AP	750	KVR	10
DELHI	900	SANSKAR	20
TS	800	FAZINA	15

In [15]:
```
grp=df.groupby("CITY")
grp.nth(1)
```

Out[15]:

	CID	CNAME	CITY	EXP
2	300	LAXMI	TS	10
3	400	ROSSUM	AP	20
5	600	VINAY	DELHI	15

In [16]:
```
print(df)
```

```
   CID       CNAME   CITY  EXP
0  100       SUSMA     TS   10
1  200    SAGATIKA     AP   15
2  300       LAXMI     TS   10
3  400      ROSSUM     AP   20
4  500      NARESH  DELHI   10
5  600       VINAY  DELHI   15
6  700  TEJASWANI     AP   15
7  800      FAZINA     TS   15
8  900     SANSKAR  DELHI   20
9  750         KVR     AP   10
```

In [17]:
```
df["EXP"]
```

Out[17]:
```
0    10
1    15
2    10
3    20
4    10
5    15
6    15
7    15
8    20
9    10
Name: EXP, dtype: int64
```

In [18]:
```
df["EXP"].mean()
```

Out[18]:
```
14.0
```

In [19]:
```
df["EXP"].max()
```

Out[19]:
```
20
```

In [20]:
```
df["EXP"].min()
```

Out[20]:
```
10
```

In [21]:
```
df["EXP"].median()
```

Out[21]:
```
15.0
```

In [22]:
```
df["EXP"].mode()
```

Out[22]:
```
0    10
1    15
Name: EXP, dtype: int64
```

In [23]:
```
df["EXP"].var()
```

Out[23]:
```
15.555555555555555
```

In [24]:
```
df["EXP"].std()
```

Out[24]:
```
3.9440531887330774
```

In [25]:
```
#Statistical Functions
```

In [26]:
```
data={'Maths':[90,85,98,80,55,78],'Science':[92,87,59,64,87,96],'English':[95,
print(data,type(data))
```

```
{'Maths': [90, 85, 98, 80, 55, 78], 'Science': [92, 87, 59, 64, 87, 96], 'Eng
lish': [95, 94, 84, 75, 67, 65]} <class 'dict'>
```

In [27]:
```
df=pd.DataFrame(data)
print(df)
```

```
   Maths  Science  English
0     90       92       95
1     85       87       94
2     98       59       84
3     80       64       75
4     55       87       67
5     78       96       65
```

In [28]:
```
df.sum()
```

Out[28]:
```
Maths      486
Science    485
English    480
dtype: int64
```

In [29]:
```
s=df["Maths"].sum()
print("Sum of Maths=",s)
```

```
Sum of Maths= 486
```

In [31]:
```
df=pd.DataFrame(data)
print(df)
df.count()
```

```
   Maths  Science  English
0     90       92       95
1     85       87       94
2     98       59       84
3     80       64       75
4     55       87       67
5     78       96       65
```

Out[31]:
```
Maths      6
Science    6
English    6
dtype: int64
```

In [32]:
```
df.max()
```

Out[32]:
```
Maths      98
Science    96
English    95
dtype: int64
```

In [33]:
```
df.min()
```

Out[33]:
```
Maths      55
Science    59
English    65
dtype: int64
```

In [34]:
```
df.mean()
```

Out[34]:
```
Maths      81.000000
Science    80.833333
English    80.000000
dtype: float64
```

In [35]:
```
df.median()
```

Out[35]:
```
Maths      82.5
Science    87.0
English    79.5
dtype: float64
```

In [36]:
```
df.mode()
```

Out[36]:

	Maths	Science	English
0	55	87.0	65
1	78	NaN	67
2	80	NaN	75
3	85	NaN	84
4	90	NaN	94
5	98	NaN	95

In [37]:
```
df.std()
```

Out[37]:
```
Maths      14.642404
Science    15.432649
English    13.084342
dtype: float64
```

In [38]:
```
df.var()
```

Out[38]:
```
Maths      214.400000
Science    238.166667
English    171.200000
dtype: float64
```

In [ ]:
```

```