ADVANCE OPERATIONS ON DATAFRAMES

PROGRAMS

#Python Program to create the dataframe with following values

	Name of Employee	Sales	Quarter	State
0	Mohak	1000	1	Rajasthan
1	Vijay	300	1	Panjab
2	Tapasi	400	1	Gujarat
3	Mansi	500	1	Goa
4	Bipin	800	1	Rajasthan
5	Mohak	1000	2	Gujarat
6	Vijay	500	2	Panjab
7	Tapasi	700	2	Gujarat
8	Mansi	50	2	Rajasthan
9	Bipin	60	2	Rajasthan
10	Mohak	1000	3	Rajasthan
11	Vijay	900	3	Panjab
12	Tapasi	750	3	Gujarat
13	Mansi	200	3	Goa
14	Bipin	300	3	Gujarat
15	Mohak	1000	4	Panjab
16	Vijay	900	4	Panjab
17	Tapasi	250	4	Gujarat
18	Mansi	750	4	Goa
19	Bipin	50	4	Rajasthan

```
from pandas import DataFrame
Employees = {'Name of Employee':
['Mohak','Vijay','Tapasi','Mansi','Bipin','Mohak','Vijay','Tapasi','Mansi','Bipin','Mohak','Vijay','Tapasi','Mansi','Bipin'],
            'Sales':
[1000,300,400,500,800,1000,500,700,50,60,1000,900,750,200,300,1000,900,250,750,50],
            'Quarter': [1,1,1,1,1,2,2,2,2,2,3,3,3,3,3,4,4,4,4,4],
            'State':
['Rajasthan','Panjab','Gujarat','Goa','Rajasthan','Gujarat','Panjab','Gujarat','Rajasthan','Rajasthan','Rajasthan','Panjab','Gujarat','Goa','Gujarat','Panjab','Panjab','Gujarat','Goa','Rajasthan']
          }
df = DataFrame(Employees, columns= ['Name of Employee', 'Sales','Quarter','State'])
print (df)
```

#Find total sales per employee in above dataframe

```
from pandas import DataFrame
Employees = {'Name of Employee':
['Mohak','Vijay','Tapasi','Mansi','Bipin','Mohak','Vijay','Tapasi','Mansi','Bipin','Mohak','Vijay','Tapasi','Mansi',
'Bipin','Mohak','Vijay','Tapasi','Mansi','Bipin'],
            'Sales':
[1000,300,400,500,800,1000,500,700,50,60,1000,900,750,200,300,1000,900,250,750,50],
            'Quarter': [1,1,1,1,1,2,2,2,2,2,3,3,3,3,3,4,4,4,4,4],
            'State':
['Rajasthan','Panjab','Gujarat','Goa','Rajasthan','Gujarat','Panjab','Gujarat','Rajasthan','Rajasthan','Rajasth
an','Panjab','Gujarat','Goa','Gujarat','Panjab','Panjab','Gujarat','Goa','Rajasthan']
          }
df = DataFrame(Employees, columns= ['Name of Employee', 'Sales','Quarter','State'])
print (df)
pivot = df.pivot_table(index=['Name of Employee'], values=['Sales'], aggfunc='sum')
print (pivot)
```

OUTPUT

	Sales
Name of Employee	
Bipin	**1210**
Mansi	**1500**
Mohak	**4000**
Tapasi	**2100**
Vijay	**2600**

#Find total sales by state in above dataframe

```
from pandas import DataFrame
Employees = {'Name of Employee':
['Mohak','Vijay','Tapasi','Mansi','Bipin','Mohak','Vijay','Tapasi','Mansi','Bipin','Mohak','Vijay','Tapasi','Mansi',
'Bipin','Mohak','Vijay','Tapasi','Mansi','Bipin'],
            'Sales':
[1000,300,400,500,800,1000,500,700,50,60,1000,900,750,200,300,1000,900,250,750,50],
            'Quarter': [1,1,1,1,1,2,2,2,2,2,3,3,3,3,3,4,4,4,4,4],
            'State':
['Rajasthan','Panjab','Gujarat','Goa','Rajasthan','Gujarat','Panjab','Gujarat','Rajasthan','Rajasthan','Rajasth
an','Panjab','Gujarat','Goa','Gujarat','Panjab','Panjab','Gujarat','Goa','Rajasthan']
          }
df = DataFrame(Employees, columns= ['Name of Employee', 'Sales','Quarter','State'])
print (df)
pivot = df.pivot_table(index=['State'], values=['Sales'], aggfunc='sum')
print (pivot)
```

OUTPUT

	Sales
State	
Goa	**1450**
Gujarat	**3400**
Panjab	**3600**
Rajasthan	**2960**

#Find total sales by both employee& state in above dataframe

```
from pandas import DataFrame
Employees = {'Name of Employee':
['Mohak','Vijay','Tapasi','Mansi','Bipin','Mohak','Vijay','Tapasi','Mansi','Bipin','Mohak','Vijay','Tapasi','Mansi',
'Bipin','Mohak','Vijay','Tapasi','Mansi','Bipin'],
            'Sales':
[1000,300,400,500,800,1000,500,700,50,60,1000,900,750,200,300,1000,900,250,750,50],
            'Quarter': [1,1,1,1,1,2,2,2,2,2,3,3,3,3,3,4,4,4,4,4],
            'State':
['Rajasthan','Panjab','Gujarat','Goa','Rajasthan','Gujarat','Panjab','Gujarat','Rajasthan','Rajasthan','Rajasth
an','Panjab','Gujarat','Goa','Gujarat','Panjab','Panjab','Gujarat','Goa','Rajasthan']
           }
df = DataFrame(Employees, columns= ['Name of Employee', 'Sales','Quarter','State'])
print (df)
pivot = df.pivot_table(index=['Name of Employee','State'], values=['Sales'], aggfunc='sum')
print (pivot)
```

OUTPUT

		Sales
Name of Employee	**State**	
Bipin	**Gujarat**	**300**
	Rajasthan	**910**
Mansi	**Goa**	**1450**
	Rajasthan	**50**
Mohak	**Gujarat**	**1000**
	Panjab	**1000**
	Rajasthan	**2000**
Tapasi	**Gujarat**	**2100**
Vijay	**Panjab**	**2600**

#Find Max individual sale by State in above dataframe

```
from pandas import DataFrame
Employees = {'Name of Employee':
['Mohak','Vijay','Tapasi','Mansi','Bipin','Mohak','Vijay','Tapasi','Mansi','Bipin','Mohak','Vijay','Tapasi','Mansi',
'Bipin','Mohak','Vijay','Tapasi','Mansi','Bipin'],
            'Sales':
[1000,300,400,500,800,1000,500,700,50,60,1000,900,750,200,300,1000,900,250,750,50],
            'Quarter': [1,1,1,1,1,2,2,2,2,2,3,3,3,3,3,4,4,4,4,4],
            'State':
['Rajasthan','Panjab','Gujarat','Goa','Rajasthan','Gujarat','Panjab','Gujarat','Rajasthan','Rajasthan','Rajasth
an','Panjab','Gujarat','Goa','Gujarat','Panjab','Panjab','Gujarat','Goa','Rajasthan']
           }
df = DataFrame(Employees, columns= ['Name of Employee', 'Sales','Quarter','State'])
print (df)
pivot = df.pivot_table(index=['State'], values=['Sales'], aggfunc='max')
print (pivot)
```

OUTPUT

	Sales
State	
Goa	**750**
Gujarat	**1000**
Panjab	**1000**
Rajasthan	**1000**

#Find Mean, median and min sales by State in above dataframe

```
from pandas import DataFrame
Employees = {'Name of Employee':
['Mohak','Vijay','Tapasi','Mansi','Bipin','Mohak','Vijay','Tapasi','Mansi','Bipin','Mohak','Vijay','Tapasi','Mansi',
'Bipin','Mohak','Vijay','Tapasi','Mansi','Bipin'],
            'Sales':
[1000,300,400,500,800,1000,500,700,50,60,1000,900,750,200,300,1000,900,250,750,50],
            'Quarter': [1,1,1,1,1,2,2,2,2,2,3,3,3,3,3,4,4,4,4,4],
            'State':
['Rajasthan','Panjab','Gujarat','Goa','Rajasthan','Gujarat','Panjab','Gujarat','Rajasthan','Rajasthan','Rajasth
an','Panjab','Gujarat','Goa','Gujarat','Panjab','Panjab','Gujarat','Goa','Rajasthan']
           }
df = DataFrame(Employees, columns= ['Name of Employee', 'Sales','Quarter','State'])
print (df)
pivot = df.pivot_table(index=['State'], values=['Sales'], aggfunc={'median','mean','min'})
print (pivot)
```

OUTPUT

	Sales		
	mean	**median**	**min**
State			
Goa	**483.333333**	**500.0**	**200.0**
Gujarat	**566.666667**	**550.0**	**250.0**
Panjab	**720.000000**	**900.0**	**300.0**
Rajasthan	**493.333333**	**430.0**	**50.0**

#Python Program to create the dataframe with following values

	name	year	score	catches
0	Mohak	2012	10	2
1	Rajesh	2012	22	2
2	Freya	2013	11	3
3	Aditya	2014	32	3
4	Anika	2014	23	3

```
import pandas as pd
data = {'name': ['Mohak', 'Rajesh', 'Freya', 'Aditya', 'Anika'],
        'year': [2012, 2012, 2013, 2014, 2014],
        'score': [10, 22, 11, 32, 23],
        'catches': [2, 2, 3, 3, 3]}
df = pd.DataFrame(data, columns= ['name', 'year','score','catches'])
print(df)
```

#Sort the dataframe’s rows by score, in descending order

```
import pandas as pd
data = {'name': ['Mohak', 'Rajesh', 'Freya', 'Aditya', 'Anika'],
        'year': [2012, 2012, 2013, 2014, 2014],
        'score': [10, 22, 11, 32, 23],
        'catches': [2, 2, 3, 3, 3]}
df = pd.DataFrame(data, columns= ['name', 'year','score','catches'])
print(df)
r=df.sort_values(by='score', ascending=False)
print(r)
```

OUTPUT

	name	year	score	catches
3	Aditya	2014	32	3
4	Anika	2014	23	3
1	Rajesh	2012	22	2
2	Freya	2013	11	3
0	Mohak	2012	10	2

#Sort the dataframe’s rows by catches and then by score, in ascending order/sort by multiple columns

```
import pandas as pd
data = {'name': ['Mohak', 'Rajesh', 'Freya', 'Aditya', 'Anika'],
        'year': [2012, 2012, 2013, 2014, 2014],
        'score': [10, 22, 11, 32, 23],
        'catches': [2, 2, 3, 3, 3]}
df = pd.DataFrame(data, columns= ['name', 'year','score','catches'])
print(df)
r=df.sort_values(by=['catches', 'score'])
print(r)
```

OUTPUT

	name	year	score	catches
0	Mohak	2012	10	2
1	Rajesh	2012	22	2
2	Freya	2013	11	3
4	Anika	2014	23	3
3	Aditya	2014	32	3

#Sort the dataframe’s rows using index

```
import pandas as pd
data = {'name': ['Mohak', 'Rajesh', 'Freya', 'Aditya', 'Anika'],
        'year': [2012, 2012, 2013, 2014, 2014],
        'score': [10, 22, 11, 32, 23],
        'catches': [2, 2, 3, 3, 3]}
df = pd.DataFrame(data, columns= ['name', 'year','score','catches'],index=[4,5,3,2,1])
print(df)
r=df.sort_index()
print(r)
```

OUTPUT

	name	year	score	catches
1	Anika	2014	23	3
2	Aditya	2014	32	3
3	Freya	2013	11	3
4	Mohak	2012	10	2
5	Rajesh	2012	22	2

#Sort the dataframe’s rows descending of index value

```
import pandas as pd
data = {'name': ['Mohak', 'Rajesh', 'Freya', 'Aditya', 'Anika'],
        'year': [2012, 2012, 2013, 2014, 2014],
        'score': [10, 22, 11, 32, 23],
        'catches': [2, 2, 3, 3, 3]}
df = pd.DataFrame(data, columns= ['name', 'year','score','catches'],index=[4,5,3,2,1])
print(df)
r=df.sort_index(ascending=False)
print(r)
```

OUTPUT

	name	year	score	catches
5	Rajesh	2012	22	2
4	Mohak	2012	10	2
3	Freya	2013	11	3
2	Aditya	2014	32	3
1	Anika	2014	23	3