Group By

In [1]:
```
import pandas as pd
```

In [3]:
```
df = pd.read_csv("dataset//weather_by_cities_group.csv")
df
```

Out[3]:

	day	city	temperature	windspeed	event
0	1/1/2017	new york	32	6	Rain
1	1/2/2017	new york	36	7	Sunny
2	1/3/2017	new york	28	12	Snow
3	1/4/2017	new york	33	7	Sunny
4	1/1/2017	mumbai	90	5	Sunny
5	1/2/2017	mumbai	85	12	Fog
6	1/3/2017	mumbai	87	15	Fog
7	1/4/2017	mumbai	92	5	Rain
8	1/1/2017	paris	45	20	Sunny
9	1/2/2017	paris	50	13	Cloudy
10	1/3/2017	paris	54	8	Cloudy
11	1/4/2017	paris	42	10	Cloudy

In [4]:
```
g = df.groupby("city")
g
```

Out[4]:
```
<pandas.core.groupby.generic.DataFrameGroupBy object at 0x000002B04A955C40>
```

In [6]:
```
for city,data in g:
    #print("\n")
    print(city)
    #print("\n")
    print(data)
```

```
mumbai
        day    city  temperature  windspeed  event
4  1/1/2017  mumbai           90          5  Sunny
5  1/2/2017  mumbai           85         12    Fog
6  1/3/2017  mumbai           87         15    Fog
7  1/4/2017  mumbai           92          5   Rain
new york
        day      city  temperature  windspeed  event
0  1/1/2017  new york           32          6   Rain
1  1/2/2017  new york           36          7  Sunny
2  1/3/2017  new york           28         12   Snow
3  1/4/2017  new york           33          7  Sunny
paris
         day   city  temperature  windspeed   event
8   1/1/2017  paris           45         20   Sunny
9   1/2/2017  paris           50         13  Cloudy
10  1/3/2017  paris           54          8  Cloudy
11  1/4/2017  paris           42         10  Cloudy
```

In [7]:
```
g1 = df.groupby("event")
g1
```

Out[7]:
```
<pandas.core.groupby.generic.DataFrameGroupBy object at 0x000002B04A955970>
```

In [8]:
```
for event,data in g1:
    print(event)
    print(data)
```

```
Cloudy
         day   city  temperature  windspeed   event
9   1/2/2017  paris           50         13  Cloudy
10  1/3/2017  paris           54          8  Cloudy
11  1/4/2017  paris           42         10  Cloudy
Fog
        day    city  temperature  windspeed event
5  1/2/2017  mumbai           85         12   Fog
6  1/3/2017  mumbai           87         15   Fog
Rain
        day      city  temperature  windspeed event
0  1/1/2017  new york           32          6  Rain
7  1/4/2017    mumbai           92          5  Rain
Snow
        day      city  temperature  windspeed event
2  1/3/2017  new york           28         12  Snow
Sunny
        day      city  temperature  windspeed  event
1  1/2/2017  new york           36          7  Sunny
3  1/4/2017  new york           33          7  Sunny
4  1/1/2017    mumbai           90          5  Sunny
8  1/1/2017     paris           45         20  Sunny
```

In [9]:
```
g1.get_group("Sunny")
```

Out[9]:

	day	city	temperature	windspeed	event
1	1/2/2017	new york	36	7	Sunny
3	1/4/2017	new york	33	7	Sunny
4	1/1/2017	mumbai	90	5	Sunny
8	1/1/2017	paris	45	20	Sunny

In [10]:
```
g1.min()
```

Out[10]:

event	day	city	temperature	windspeed
Cloudy	1/2/2017	paris	42	8
Fog	1/2/2017	mumbai	85	12
Rain	1/1/2017	mumbai	32	5
Snow	1/3/2017	new york	28	12
Sunny	1/1/2017	mumbai	33	5

In [11]:
```
g1.max()
```

Out[11]:

event	day	city	temperature	windspeed
Cloudy	1/4/2017	paris	54	13
Fog	1/3/2017	mumbai	87	15
Rain	1/4/2017	new york	92	6
Snow	1/3/2017	new york	28	12
Sunny	1/4/2017	paris	90	20

In [12]:
```
g1.count()
```

Out[12]:

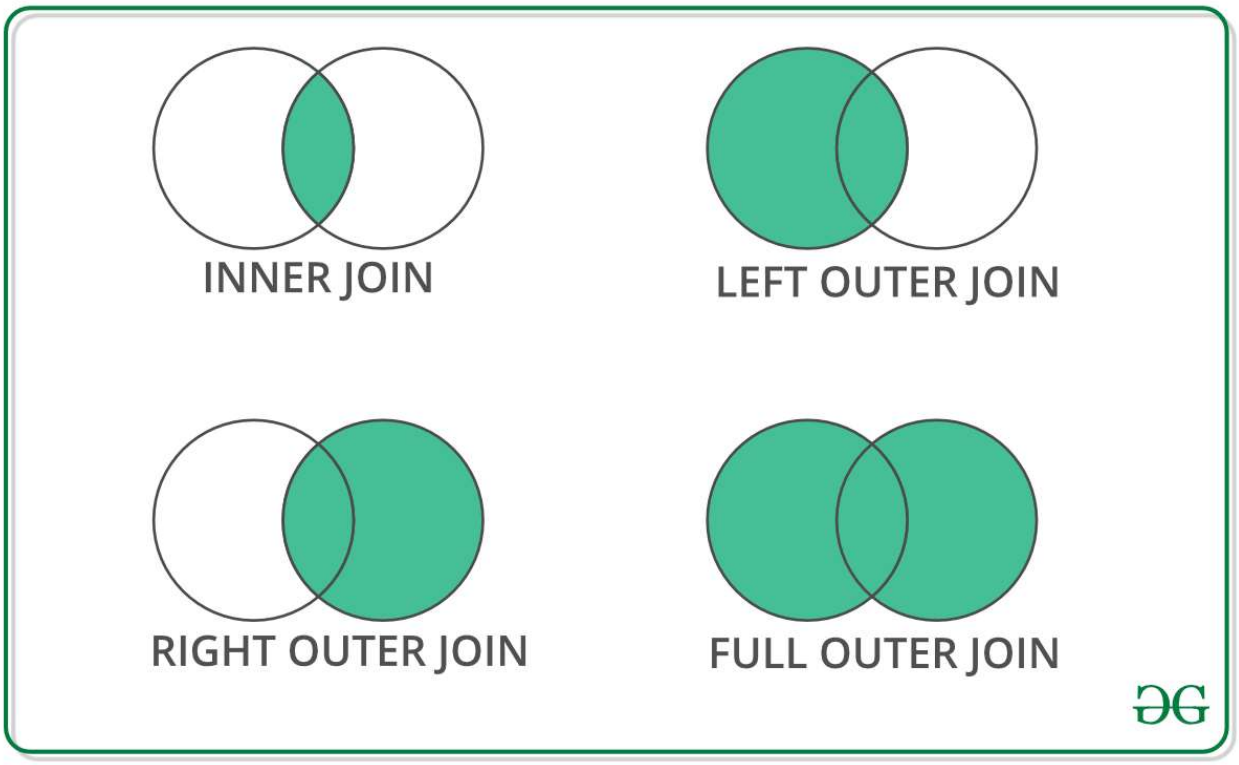
event	day	city	temperature	windspeed
Cloudy	3	3	3	3
Fog	2	2	2	2
Rain	2	2	2	2
Snow	1	1	1	1
Sunny	4	4	4	4

In [13]:
```
g1.mean()
```

Out[13]:

event	temperature	windspeed
Cloudy	48.666667	10.333333
Fog	86.000000	13.500000
Rain	62.000000	5.500000
Snow	28.000000	12.000000
Sunny	51.000000	9.750000

Join



In [14]:
```
t = {
    "Courses": ["Python","C","C++"],
    "Fee": [25000,20000,30000],
    "Duration": ["30days","25days","45days"]
}
```

In [15]:
```
d1 = pd.DataFrame(t, index=["m1","m2","m3"])
```

In [16]:
```
d1
```

Out[16]:

	Courses	Fee	Duration
m1	Python	25000	30days
m2	C	20000	25days
m3	C++	30000	45days

In [17]:
```
t2 = {
    "Courses": ["Django","Excel","C++"],
    "Fee": [23000,20000,10000]

}
d2= pd.DataFrame(t2,index=["m2","m3","m4"])
d2
```

Out[17]:

	Courses	Fee
m2	Django	23000
m3	Excel	20000
m4	C++	10000

inner join

In [19]:
```
d1.join(d2,lsuffix="_left",rsuffix="_right",how="inner")
```

Out[19]:

	Courses_left	Fee_left	Duration	Courses_right	Fee_right
m2	C	20000	25days	Django	23000
m3	C++	30000	45days	Excel	20000

left Join

In [21]:
```
d1.join(d2,lsuffix="_dataframe1",rsuffix="_dataframe2",how="left")
```

Out[21]:

	Courses_dataframe1	Fee_dataframe1	Duration	Courses_dataframe2	Fee_dataframe2
m1	Python	25000	30days	NaN	NaN
m2	C	20000	25days	Django	23000.0
m3	C++	30000	45days	Excel	20000.0

Right Join

In [22]:
```
d1.join(d2,lsuffix="_dataframe1",rsuffix="_dataframe2",how="right")
```

Out[22]:

	Courses_dataframe1	Fee_dataframe1	Duration	Courses_dataframe2	Fee_dataframe2
m2	C	20000.0	25days	Django	23000
m3	C++	30000.0	45days	Excel	20000
m4	NaN	NaN	NaN	C++	10000

Full outer Join

In [23]:
```
d1.join(d2,lsuffix="_dataframe1",rsuffix="_dataframe2",how="outer")
```

Out[23]:

	Courses_dataframe1	Fee_dataframe1	Duration	Courses_dataframe2	Fee_dataframe2
m1	Python	25000.0	30days	NaN	NaN
m2	C	20000.0	25days	Django	23000.0
m3	C++	30000.0	45days	Excel	20000.0
m4	NaN	NaN	NaN	C++	10000.0

Merge

In [24]:
```
d1
```

Out[24]:

	Courses	Fee	Duration
m1	Python	25000	30days
m2	C	20000	25days
m3	C++	30000	45days

In [25]:
```
d2
```

Out[25]:

	Courses	Fee
m2	Django	23000
m3	Excel	20000
m4	C++	10000

In [26]:
```
pd.merge(d1,d2,on="Courses",how="inner")
```

Out[26]:

	Courses	Fee_x	Duration	Fee_y
0	C++	30000	45days	10000

In [27]:
```
pd.merge(d1,d2,on="Courses",how="left")
```

Out[27]:

	Courses	Fee_x	Duration	Fee_y
0	Python	25000	30days	NaN
1	C	20000	25days	NaN
2	C++	30000	45days	10000.0

In [28]:
```
pd.merge(d1,d2,on="Courses",how="right")
```

Out[28]:

	Courses	Fee_x	Duration	Fee_y
0	Django	NaN	NaN	23000
1	Excel	NaN	NaN	20000
2	C++	30000.0	45days	10000

In [29]:
```
pd.merge(d1,d2,on="Courses",how="outer")
```

Out[29]:

	Courses	Fee_x	Duration	Fee_y
0	Python	25000.0	30days	NaN
1	C	20000.0	25days	NaN
2	C++	30000.0	45days	10000.0
3	Django	NaN	NaN	23000.0
4	Excel	NaN	NaN	20000.0

In [31]:
```
pd.merge(d1,d2,on="Courses",how="outer",suffixes=('_dataframe1','_dataframe2'))
```

Out[31]:

	Courses	Fee_dataframe1	Duration	Fee_dataframe2
0	Python	25000.0	30days	NaN
1	C	20000.0	25days	NaN
2	C++	30000.0	45days	10000.0
3	Django	NaN	NaN	23000.0
4	Excel	NaN	NaN	20000.0

In [32]:
```
pd.merge(d1,d2,on="Courses",how="outer",suffixes=('_dataframe1','_dataframe2'),indicator=Tru
```

Out[32]:

	Courses	Fee_dataframe1	Duration	Fee_dataframe2	_merge
0	Python	25000.0	30days	NaN	left_only
1	C	20000.0	25days	NaN	left_only
2	C++	30000.0	45days	10000.0	both
3	Django	NaN	NaN	23000.0	right_only
4	Excel	NaN	NaN	20000.0	right_only

Concat

In [38]:
```
df1 = pd.DataFrame({
    "city": ["delhi","mumbai","pune"],
    "temp": [45,35,30],
    "humidity": [65,70,55]
})
print(df1)

df2 = pd.DataFrame({
    "city": ["New York","mumbai","pune"],
    "temp": [50,33,31],
    "humidity": [63,73,53]
})
print(df2)
df3 = pd.DataFrame({
    "city": ["New York","mumbai","Banlore"],
    "temp": [50,33,31],
    "humidity": [63,73,53]
})
print(df3)
```

```
     city  temp  humidity
0   delhi    45        65
1  mumbai    35        70
2    pune    30        55
       city  temp  humidity
0  New York    50        63
1    mumbai    33        73
2      pune    31        53
       city  temp  humidity
0  New York    50        63
1    mumbai    33        73
2   Banlore    31        53
```

In [40]:
```
df_concat = pd.concat([df1,df2,df3])
df_concat
```

Out[40]:

	city	temp	humidity
0	delhi	45	65
1	mumbai	35	70
2	pune	30	55
0	New York	50	63
1	mumbai	33	73
2	pune	31	53
0	New York	50	63
1	mumbai	33	73
2	Banlore	31	53

In [41]:
```
df_concat = pd.concat([df1,df2,df3],ignore_index=True)
df_concat
```

Out[41]:

	city	temp	humidity
0	delhi	45	65
1	mumbai	35	70
2	pune	30	55
3	New York	50	63
4	mumbai	33	73
5	pune	31	53
6	New York	50	63
7	mumbai	33	73
8	Banlore	31	53

In [43]:
```
df_concat = pd.concat([df1,df2,df3],keys=["India","US","In-US"])
df_concat
```

Out[43]:

		city	temp	humidity
India	**0**	delhi	45	65
	1	mumbai	35	70
	2	pune	30	55
US	**0**	New York	50	63
	1	mumbai	33	73
	2	pune	31	53
In-US	**0**	New York	50	63
	1	mumbai	33	73
	2	Banlore	31	53

In [44]:
```
df_concat
```

Out[44]:

		city	temp	humidity
India	**0**	delhi	45	65
	1	mumbai	35	70
	2	pune	30	55
US	**0**	New York	50	63
	1	mumbai	33	73
	2	pune	31	53
In-US	**0**	New York	50	63
	1	mumbai	33	73
	2	Banlore	31	53

In [46]:
```
df_concat = pd.concat([df1,df2,df3],keys=["India","US","In-US"],axis=1)
df_concat
```

Out[46]:

	India			US			In-US		
	city	temp	humidity	city	temp	humidity	city	temp	humidity
0	delhi	45	65	New York	50	63	New York	50	63
1	mumbai	35	70	mumbai	33	73	mumbai	33	73
2	pune	30	55	pune	31	53	Banlore	31	53

In [ ]: