In [1]:
```
# Group By

import pandas as pd
```

In [2]:
```
df = pd.read_csv("weather_by_cities_group.csv")
df
```

Out[2]:

	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 [3]:
```
df_grpby = df.groupby("city")
df_grpby
```

Out[3]:
```
<pandas.core.groupby.generic.DataFrameGroupBy object at 0x0000019EF7A541A0>
```

In [8]:
```
for city , data in df_grpby:
    print(city)
    print("=" *60)
    #print("\n")
    print(data)
    print("=" *60)
```

```
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 [9]:
```
# get group data

df_grpby.get_group("paris")
```

Out[9]:

	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 [10]:
```
# max
df_grpby.max()
```

Out[10]:

city	day	temperature	windspeed	event
mumbai	1/4/2017	92	15	Sunny
new york	1/4/2017	36	12	Sunny
paris	1/4/2017	54	20	Sunny

In [11]:
```
# min
df_grpby.min()
```

Out[11]:

city	day	temperature	windspeed	event
mumbai	1/1/2017	85	5	Fog
new york	1/1/2017	28	6	Rain
paris	1/1/2017	42	8	Cloudy

In [12]:
```
# count
df_grpby.count()
```

Out[12]:

city	day	temperature	windspeed	event
mumbai	4	4	4	4
new york	4	4	4	4
paris	4	4	4	4

In [24]:
```
df_grpby_event = df.groupby("event")
df_grpby_event
```

Out[24]: <pandas.core.groupby.generic.DataFrameGroupBy object at 0x0000019E9F26FED0>

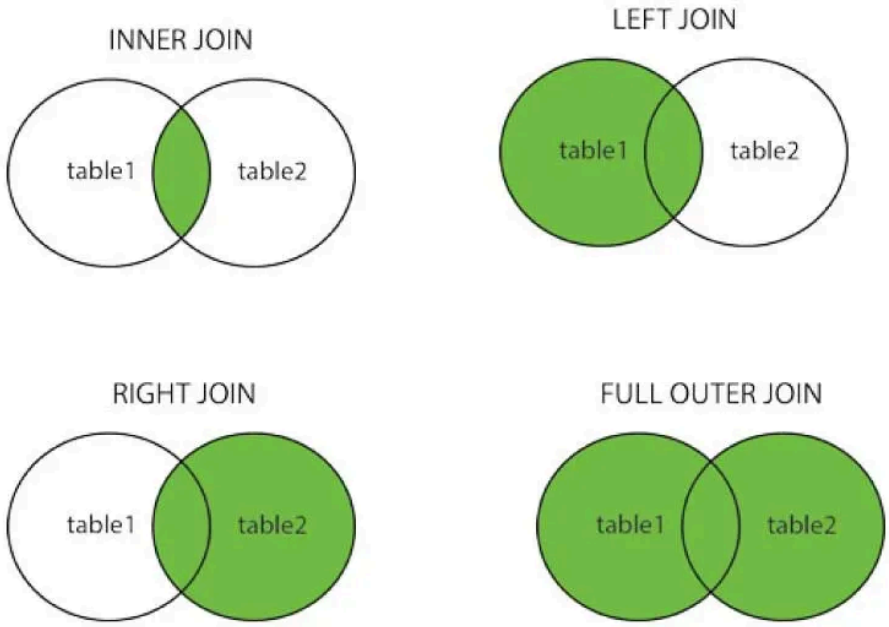
In [27]:
```
for event , data in df_grpby_event:
    print(event)
    print(data)
```

```
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 [31]:
```
for event , data in df_grpby_event:
    print(event)
    print(data["windspeed"].mean())
```

```
Cloudy
10.333333333333334
Fog
13.5
Rain
5.5
Snow
12.0
Sunny
9.75
```

Join



In [33]:
```
t1= {
    "Courses" : ["Python","C","C++"],
    "Fee" : [20000,25000,30000],
    "Duration": ["30days","40days","50days"]
}

df1 = pd.DataFrame(t1,index =["m1","m2","m3"])
df1
```

Out[33]:

	Courses	Fee	Duration
m1	Python	20000	30days
m2	C	25000	40days
m3	C++	30000	50days

In [34]:
```
t2= {
    "Courses" : ["Django","Power BI","Excel"],
    "Discount" : [2000,3000,4500]
```

```
}

df2 = pd.DataFrame(t2,index =["m2","m3","m4"])
df2
```

Out[34]:

	Courses	Discount
m2	Django	2000
m3	Power BI	3000
m4	Excel	4500

In [37]:
```
# Inner join

df_com = df1.join(df2,lsuffix="_dataframe1",rsuffix="_dataframe2",how="inner")
df_com
```

Out[37]:

	Courses_dataframe1	Fee	Duration	Courses_dataframe2	Discount
m2	C	25000	40days	Django	2000
m3	C++	30000	50days	Power BI	3000

In [38]:
```
# left join

df_com = df1.join(df2,lsuffix="_dataframe1",rsuffix="_dataframe2",how="left")
df_com
```

Out[38]:

	Courses_dataframe1	Fee	Duration	Courses_dataframe2	Discount
m1	Python	20000	30days	NaN	NaN
m2	C	25000	40days	Django	2000.0
m3	C++	30000	50days	Power BI	3000.0

In [39]:
```
# right join

df_com = df1.join(df2,lsuffix="_dataframe1",rsuffix="_dataframe2",how="right")
df_com
```

Out[39]:

	Courses_dataframe1	Fee	Duration	Courses_dataframe2	Discount
m2	C	25000.0	40days	Django	2000
m3	C++	30000.0	50days	Power BI	3000
m4	NaN	NaN	NaN	Excel	4500

In [40]:
```
# outer join

df_com = df1.join(df2,lsuffix="_dataframe1",rsuffix="_dataframe2",how="outer")
df_com
```

Out[40]:

	Courses_dataframe1	Fee	Duration	Courses_dataframe2	Discount
m1	Python	20000.0	30days	NaN	NaN
m2	C	25000.0	40days	Django	2000.0
m3	C++	30000.0	50days	Power BI	3000.0
m4	NaN	NaN	NaN	Excel	4500.0

Merge

In [41]:
```
df3 = pd.DataFrame({
    "city" : ["delhi","mumbai","pune"],
    "temp": [40,35,30]
})
df3
```

Out[41]:

	city	temp
0	delhi	40
1	mumbai	35
2	pune	30

In [43]:
```
df4 = pd.DataFrame({
    "city" : ["delhi","paris","new york"],
    "humidity": [50,55,60]
})
df4
```

Out[43]:

	city	humidity
0	delhi	50
1	paris	55
2	new york	60

In [44]:
```
# inner
df_merge = pd.merge(df3,df4,on="city", how = "inner")
df_merge
```

Out[44]:

	city	temp	humidity
0	delhi	40	50

In [45]:
```
# left
df_merge = pd.merge(df3,df4,on="city", how = "left")
df_merge
```

Out[45]:

	city	temp	humidity
0	delhi	40	50.0
1	mumbai	35	NaN
2	pune	30	NaN

In [46]:
```
# right
df_merge = pd.merge(df3,df4,on="city", how = "right")
df_merge
```

Out[46]:

	city	temp	humidity
0	delhi	40.0	50
1	paris	NaN	55
2	new york	NaN	60

In [48]:
```
# right
df_merge = pd.merge(df3,df4,on="city", how = "outer",indicator=True)
df_merge
```

Out[48]:

	city	temp	humidity	_merge
0	delhi	40.0	50.0	both
1	mumbai	35.0	NaN	left_only
2	new york	NaN	60.0	right_only
3	paris	NaN	55.0	right_only
4	pune	30.0	NaN	left_only

In [ ]:
```
# what is the difference between join and merge?
```