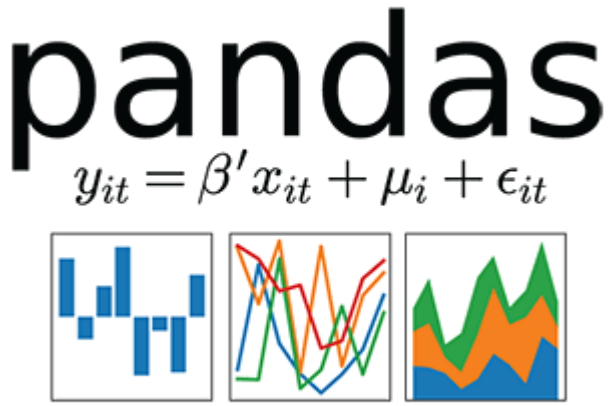


Outlines

- Pandas Groupby problem
- Print DataFrameGroupby object
- Split Apply Combine
- Find the average wind speed per city
- Access all the analytics in one shot
- Plot groupby

Pandas Groupby problem

1. Find the maximum temperature in each of the cities
2. Fine the average wind speed per city

In [15]:
```
import pandas as pd
df = pd.read_csv('D:/Data_Science/My Github/Pandas-tutorial/Document/Pandas Groupb
df
```

Out[15]:

	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 [16]:
```
g = df.groupby('city')
g
```

Out[16]:
```
<pandas.core.groupby.generic.DataFrameGroupBy object at 0x0000000008DC29D0>
```

print DataFrameGroupby object

In [18]:
```
# print DataFrameGroupby object
for city,city_df in g:
    print(city)
    print(city_df)
```

```
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 [19]:
```
# access a specific DataFrame
g.get_group('mumbai')
```

Out[19]:

	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

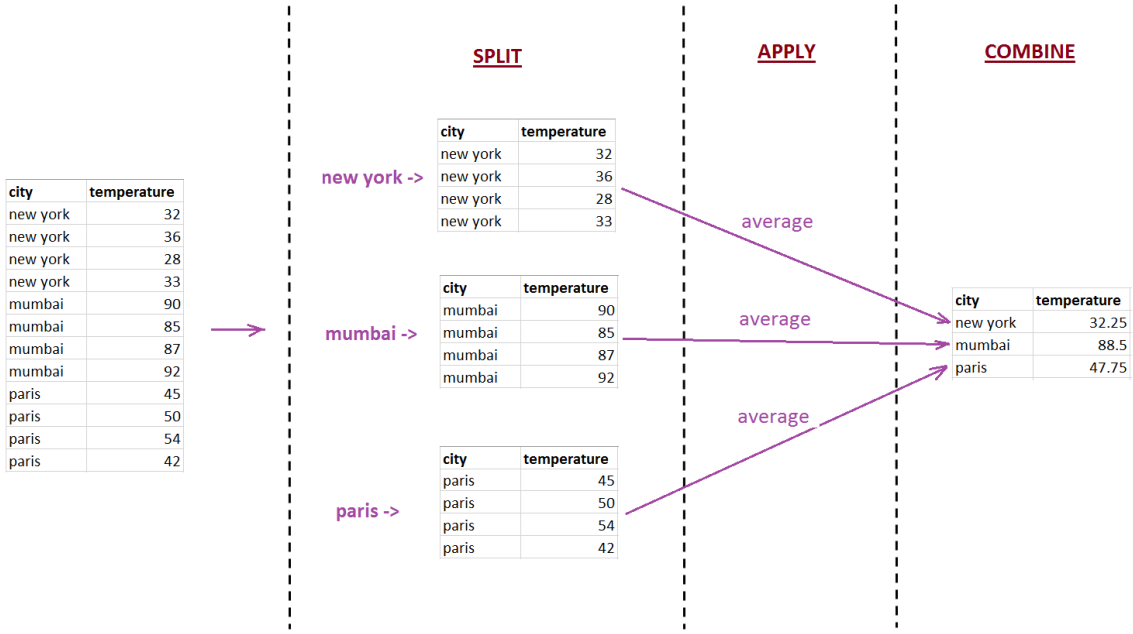
In [20]:
```
# Find the maximum temperature in each of the cities
g.max()
```

Out[20]:

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

Split Apply Combine

The process of dividing your data into multiple groups and then applying some analytics to get aggrigated result is called split apply combine



city	temperature
new york	32
new york	36
new york	28
new york	33
mumbai	90
mumbai	85
mumbai	87
mumbai	92
paris	45
paris	50
paris	54
paris	42

city	temperature
new york	32.25
mumbai	88.5
paris	47.75

Fine the average wind speed per city

In [21]:
```
g.mean()
```

Out[21]:

city	temperature	windspeed
mumbai	88.50	9.25
new york	32.25	8.00
paris	47.75	12.75

access all the analytics in one shot

In [22]:
```
g.describe()
```

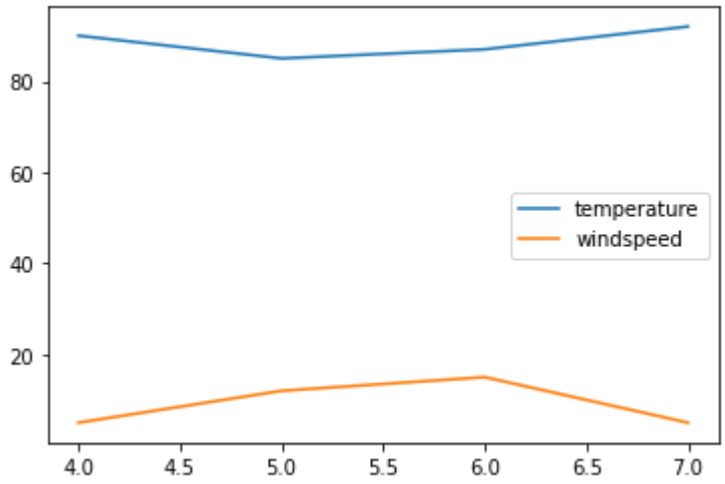
Out[22]:

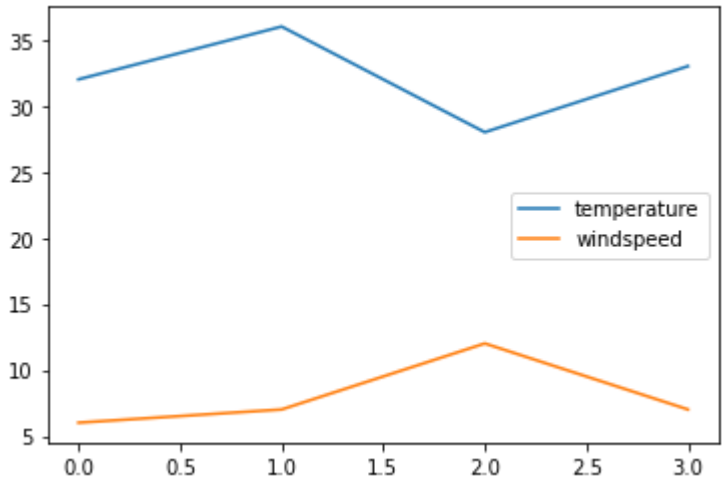
	temperature								windspeed				
	count	mean	std	min	25%	50%	75%	max	count	mean	std	min	25%
city													
mumbai	4.0	88.50	3.109126	85.0	86.50	88.5	90.50	92.0	4.0	9.25	5.057997	5.0	5.00
new york	4.0	32.25	3.304038	28.0	31.00	32.5	33.75	36.0	4.0	8.00	2.708013	6.0	6.75
paris	4.0	47.75	5.315073	42.0	44.25	47.5	51.00	54.0	4.0	12.75	5.251984	8.0	9.50

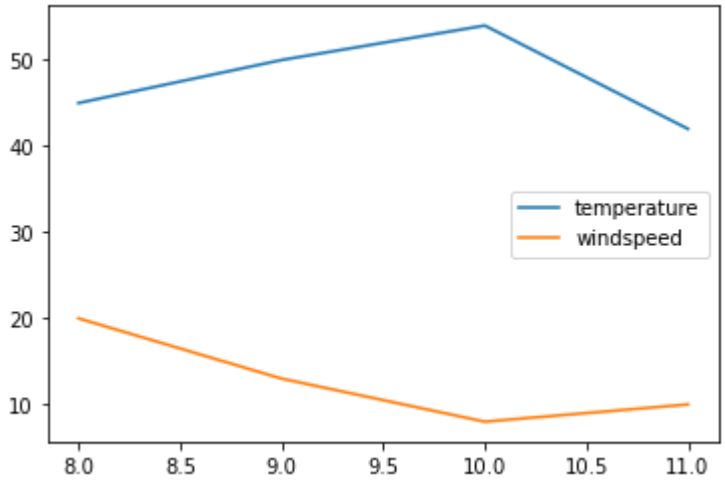
plot groupby

In [23]:
```
%matplotlib inline
g.plot()
```

Out[23]:
```
city
mumbai      AxesSubplot(0.125,0.125;0.775x0.755)
new york    AxesSubplot(0.125,0.125;0.775x0.755)
paris       AxesSubplot(0.125,0.125;0.775x0.755)
dtype: object
```







[Click here for more information about pandas groupby (https://pandas.pydata.org/pandas-

docs/stable/reference/api/pandas.DataFrame.groupby.html)

Date	Author
2021-07-28	**Ehsan Zia**