groupby and aggregate function in panda¶

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

In [2]:
```
df = pd.read_csv('C:\\Users\\Asus\\Downloads\\Flavors.csv')
df
```

Out[2]:

	Flavor	Base Flavor	Liked	Flavor Rating	Texture Rating	Total Rating
0	Mint Chocolate Chip	Vanilla	Yes	10.0	8.0	18.0
1	Chocolate	Chocolate	Yes	8.8	7.6	16.6
2	Vanilla	Vanilla	No	4.7	5.0	9.7
3	Cookie Dough	Vanilla	Yes	6.9	6.5	13.4
4	Rocky Road	Chocolate	Yes	8.2	7.0	15.2
5	Pistachio	Vanilla	No	2.3	3.4	5.7
6	Cake Batter	Vanilla	Yes	6.5	6.0	12.5
7	Neapolitan	Vanilla	No	3.8	5.0	8.8
8	Chocolte Fudge Brownie	Chocolate	Yes	8.2	7.1	15.3

In [3]:
```
df.groupby('Base Flavor')
```

Out[3]:
```
<pandas.core.groupby.generic.DataFrameGroupBy object at 0x000001E11939DE40>
```

In [17]:
```
df1=df.groupby('Base Flavor')
```

In [5]:
```
df1.mean()
```

```
C:\Users\Asus\AppData\Local\Temp\ipykernel_7568\2053335143.py:1: FutureWarning: The default value of num
eric_only in DataFrameGroupBy.mean is deprecated. In a future version, numeric_only will default to Fals
e. Either specify numeric_only or select only columns which should be valid for the function.
  df1.mean()
```

Out[5]:

Base Flavor	Flavor Rating	Texture Rating	Total Rating
Chocolate	8.4	7.233333	15.70
Vanilla	5.7	5.650000	11.35

In [6]:
```
df.groupby('Base Flavor').mean()
```

```
C:\Users\Asus\AppData\Local\Temp\ipykernel_7568\2409314591.py:1: FutureWarning: The default value of num
eric_only in DataFrameGroupBy.mean is deprecated. In a future version, numeric_only will default to Fals
e. Either specify numeric_only or select only columns which should be valid for the function.
  df.groupby('Base Flavor').mean()
```

Out[6]:

Base Flavor	Flavor Rating	Texture Rating	Total Rating
Chocolate	8.4	7.233333	15.70
Vanilla	5.7	5.650000	11.35

In [7]:
```
df.groupby('Base Flavor').count()
```

Out[7]:

Base Flavor	Flavor	Liked	Flavor Rating	Texture Rating	Total Rating
Chocolate	3	3	3	3	3
Vanilla	6	6	6	6	6

In [8]:
```
df.groupby('Base Flavor').min()
```

Out[8]:

Base Flavor	Flavor	Liked	Flavor Rating	Texture Rating	Total Rating
Chocolate	Chocolate	Yes	8.2	7.0	15.2
Vanilla	Cake Batter	No	2.3	3.4	5.7

In [9]:
```
df.groupby('Base Flavor').max()
```

Out[9]:

Base Flavor	Flavor	Liked	Flavor Rating	Texture Rating	Total Rating
Chocolate	Rocky Road	Yes	8.8	7.6	16.6
Vanilla	Vanilla	Yes	10.0	8.0	18.0

In [10]:
```
df.groupby('Base Flavor').sum()
```

```
C:\Users\Asus\AppData\Local\Temp\ipykernel_7568\1221781544.py:1: FutureWarning: The default value of num
eric_only in DataFrameGroupBy.sum is deprecated. In a future version, numeric_only will default to Fals
e. Either specify numeric_only or select only columns which should be valid for the function.
  df.groupby('Base Flavor').sum()
```

Out[10]:

Base Flavor	Flavor Rating	Texture Rating	Total Rating
Chocolate	25.2	21.7	47.1
Vanilla	34.2	33.9	68.1

In [11]:
```
f.groupby('Base Flavor').agg({'Flavor Rating': ['mean','max','count','sum'], 'Texture Rating':['mean','max
```

Out[11]:

	Flavor Rating				Texture Rating			
	mean	max	count	sum	mean	max	count	sum
Base Flavor								
Chocolate	8.4	8.8	3	25.2	7.233333	7.6	3	21.7
Vanilla	5.7	10.0	6	34.2	5.650000	8.0	6	33.9

In [12]:
```
df.groupby(['Base Flavor','Liked']).mean()
```

```
C:\Users\Asus\AppData\Local\Temp\ipykernel_7568\2854014148.py:1: FutureWarning: The default value of num
eric_only in DataFrameGroupBy.mean is deprecated. In a future version, numeric_only will default to Fals
e. Either specify numeric_only or select only columns which should be valid for the function.
  df.groupby(['Base Flavor','Liked']).mean()
```

Out[12]:

Base Flavor	Liked	Flavor Rating	Texture Rating	Total Rating
Chocolate	**Yes**	8.4	7.233333	15.700000
Vanilla	**No**	3.6	4.466667	8.066667
	Yes	7.8	6.833333	14.633333

In [13]:
```
df.groupby(['Base Flavor','Liked']).agg({'Flavor Rating': ['mean','max','min','count','sum']})
```

Out[13]:

		Flavor Rating				
		mean	max	min	count	sum
Base Flavor	Liked					
Chocolate	Yes	8.4	8.8	8.2	3	25.2
Vanilla	No	3.6	4.7	2.3	3	10.8
	Yes	7.8	10.0	6.5	3	23.4

In [14]:
```
df.groupby('Base Flavor').describe()
```

Out[14]:

	Flavor Rating								Texture Rating		...			Total Rating				
	count	mean	std	min	25%	50%	75%	max	count	mean	...	75%	max	count	mean	std	min	25
Base Flavor																		
Chocolate	3.0	8.4	0.346410	8.2	8.200	8.2	8.5	8.8	3.0	7.233333	...	7.350	7.6	3.0	15.70	0.781025	15.2	15
Vanilla	6.0	5.7	2.710719	2.3	4.025	5.6	6.8	10.0	6.0	5.650000	...	6.375	8.0	6.0	11.35	4.263684	5.7	9

2 rows × 24 columns

In [15]:
```
df.aggregate(['mean'])
```

```
C:\Users\Asus\AppData\Local\Temp\ipykernel_7568\2855992873.py:1: FutureWarning: ['Flavor', 'Base Flavo
r', 'Liked'] did not aggregate successfully. If any error is raised this will raise in a future version
of pandas. Drop these columns/ops to avoid this warning.
  df.aggregate(['mean'])
```

Out[15]:

	Flavor Rating	Texture Rating	Total Rating
mean	6.6	6.177778	12.8

In [ ]:

In [ ]:

In [ ]: